

## American Academy of Clinical Neuropsychology (AACN) 2021 consensus statement on validity assessment: Update of the 2009 AACN consensus conference statement on neuropsychological assessment of effort, response bias, and malingering

Jerry J. Sweet, Robert L. Heilbronner, Joel E. Morgan, Glenn J. Larrabee, Martin L. Rohling, Kyle B. Boone, Michael W. Kirkwood, Ryan W. Schroeder, Julie A. Suhr & Conference Participants

To cite this article: Jerry J. Sweet, Robert L. Heilbronner, Joel E. Morgan, Glenn J. Larrabee, Martin L. Rohling, Kyle B. Boone, Michael W. Kirkwood, Ryan W. Schroeder, Julie A. Suhr & Conference Participants (2021) American Academy of Clinical Neuropsychology (AACN) 2021 consensus statement on validity assessment: Update of the 2009 AACN consensus conference statement on neuropsychological assessment of effort, response bias, and malingering, *The Clinical Neuropsychologist*, 35:6, 1053-1106, DOI: [10.1080/13854046.2021.1896036](https://doi.org/10.1080/13854046.2021.1896036)

To link to this article: <https://doi.org/10.1080/13854046.2021.1896036>



Published online: 06 Apr 2021.



Submit your article to this journal [↗](#)



Article views: 7607



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 17 View citing articles [↗](#)



# American Academy of Clinical Neuropsychology (AACN) 2021 consensus statement on validity assessment: Update of the 2009 AACN consensus conference statement on neuropsychological assessment of effort, response bias, and malingering

Jerry J. Sweet<sup>a</sup>, Robert L. Heilbronner<sup>b</sup>, Joel E. Morgan<sup>c</sup>, Glenn J. Larrabee<sup>d</sup>,  
Martin L. Rohling<sup>e</sup>, Kyle B. Boone<sup>f</sup>, Michael W. Kirkwood<sup>g</sup>, Ryan W. Schroeder<sup>h</sup>,  
Julie A. Suhr<sup>i</sup> and Conference Participants<sup>#</sup>

<sup>a</sup>Department of Psychiatry & Behavioral Sciences, NorthShore University HealthSystem, Evanston, IL, USA; <sup>b</sup>Independent Practice, Chicago, IL, USA; <sup>c</sup>Independent Practice, Edison, NJ, USA; <sup>d</sup>Independent Practice, Sarasota, FL, USA; <sup>e</sup>Psychology Department, University of South Alabama, Mobile, AL, USA; <sup>f</sup>California School of Forensic Studies, Alliant International University, Los Angeles, CA, USA; <sup>g</sup>Department of Physical Medicine & Rehabilitation, University of Colorado School of Medicine and Children's Hospital Colorado, Aurora, CO, USA; <sup>h</sup>Department of Psychiatry and Behavioral Sciences, University of Kansas School of Medicine, Wichita, KS, USA; <sup>i</sup>Psychology Department, Ohio University, Athens, OH, USA

## ABSTRACT

**Objective:** Citation and download data pertaining to the 2009 AACN consensus statement on validity assessment indicated that the topic maintained high interest in subsequent years, during which key terminology evolved and relevant empirical research proliferated. With a general goal of providing current guidance to the clinical neuropsychology community regarding this important topic, the specific update goals were to: identify current key definitions of terms relevant to validity assessment; learn what experts believe should be reaffirmed from the original consensus paper, as well as new consensus points; and incorporate the latest recommendations regarding the use of validity testing, as well as current application of the term 'malingering.' **Methods:** In the spring of 2019, four of the original 2009 work group chairs and additional experts for each work group were impaneled. A total of 20 individuals shared ideas and writing drafts until reaching consensus on January 21, 2021. **Results:** Consensus was reached regarding affirmation of prior salient points that continue to garner clinical and scientific support, as well as creation of new points. The resulting consensus statement addresses definitions and differential diagnosis, performance and symptom validity assessment, and research design and statistical issues. **Conclusions/Importance:** In order to provide bases for diagnoses and interpretations, the current consensus is that all clinical and forensic evaluations must proactively address the degree to which results of neuropsychological and psychological testing are valid.

## ARTICLE HISTORY

Received 16 February 2021  
Accepted 23 February 2021  
Published online 6 April 2021

**CONTACT** Jerry J. Sweet  [jerrysweet121@gmail.com](mailto:jerrysweet121@gmail.com)  Department of Psychiatry & Behavioral Sciences, NorthShore University HealthSystem, 909 Davis St., Suite 160, Evanston, IL 60201, USA  
#Consensus Conference Statement Update Participants.

© 2021 Informa UK Limited, trading as Taylor & Francis Group

There is a strong and continually-growing evidence-based literature on which practitioners can confidently base their judgments regarding the selection and interpretation of validity measures.

## Consensus conference statement update participants

Co-Organizers: Jerry J. Sweet, Robert L. Heilbronner

*Topic Work Groups:*

**Definitions and Differential Diagnosis:** Core Writers: Robert L. Heilbronner (Chair), Ryan W. Schroeder; Internal Work Group Contributors: Shane S. Bush, Daryl E.M. Fujii, Phillip K. Martin

**Performance Validity Assessment:** Core Writers: Jerry J. Sweet (Chair), Kyle B. Boone, Michael W. Kirkwood; Internal Work Group Contributors: Michael D. Chafetz, Nathaniel W. Nelson

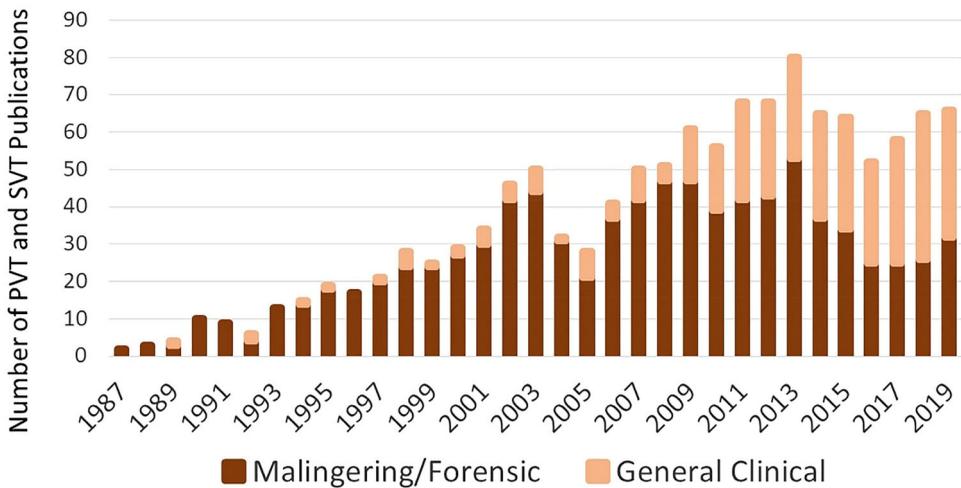
**Self-Report of Psychological and Somatic Symptoms:** Core Writers: Joel E. Morgan (Chair), Julie A. Suhr; Internal Work Group Contributors: Yossi Ben-Porath, Robert L. Denney, Roger O. Gervais

**Research Design and Statistical Issues:** Core Writers: Glenn J. Larrabee (Co-Chair) and Martin L. Rohling (Co-Chair); Internal Work Group Contributors: William B. Barr, Jeremy J. Davis, Jacobus Donders

## Introduction

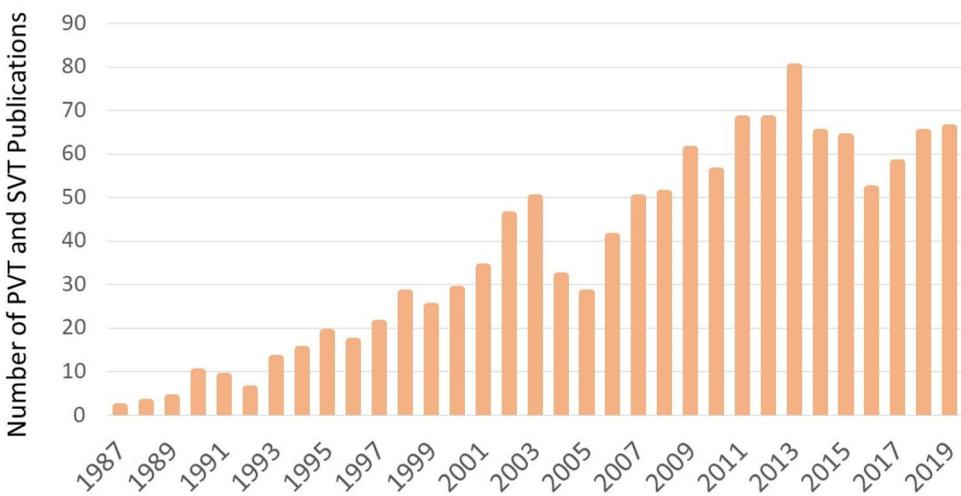
Guidance to neuropsychology's practitioners regarding validity assessment and malingering was initially addressed via a first-ever American Academy of Clinical Neuropsychology (AACN) consensus conference in 2008. Subsequently, a 36-page consensus statement was published in AACN's official journal, *The Clinical Neuropsychologist*, with the title: "American Academy of Clinical Neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering" (Heilbronner et al., 2009). This consensus statement proved especially important in providing assistance to clinical neuropsychologists as they addressed important issues of differential diagnosis in forensic and clinical cases. Even ten years post-publication, as of March 7, 2021, the original consensus statement has continued in the second place rank of most highly cited articles (<https://www.tandfonline.com/action/showMostCitedArticles?journalCode=ntcn20>), and second "most read" article (<https://www.tandfonline.com/action/showMostReadArticles?journalCode=ntcn20>) of *The Clinical Neuropsychologist*. Reflecting newer terminology and current thinking, the present update to the original 2009 consensus statement will be referred to as "the consensus statement on validity assessment."

In the years since the original consensus statement, there have been indications of an impressive rate of continued production of relevant peer-reviewed and scholarly literature on validity assessment and malingering (cf. Sweet & Guidotti Breting, 2013;



**Figure 1.** Depiction of general trends of “clinical” versus “forensic” neuropsychology-relevant publications that focus on Performance Validity Test/Symptom Validity Test (PVT/SVT) issues, identified by using multiple ‘wild card’ search terms across all peer-reviewed journals catalogued in the PsycINFO and MEDLINE databases from 1987 through 2019.

Reprinted from Suchy (2019) with permission.



**Figure 2.** Number of validity testing publications in each year from 1987 until 2019, identified by using multiple ‘wild card’ search terms across all peer-reviewed journals catalogued in the PsycINFO and MEDLINE databases.

Reprinted and modified from the original in Suchy (2019) with permission.

Suchy, 2019). In the introduction to a special issue on validity assessment in *The Clinical Neuropsychologist*, Suchy (2019) effectively depicted in graphic form the transition from neuropsychological validity assessment research that focused mostly on forensic contexts to validity assessment research that focuses with greater frequency on clinical contexts (see Figure 1), with a continued overall high rate of annual publishing covering both topics (see Figure 2).

Over the years, the breadth and depth of expert guidance on validity assessment has at times raised new issues, and caused basic terminology to evolve. Similarly, the continued applications of pre-2009 measures and the availability and applications of post-2009 measures have raised new questions, and, in some instances, have caused rethinking about past questions and answers. Moreover, forensic and clinical venues in which these issues are addressed, as well as the general sophistication of the practitioners in these venues, have evolved to varying degrees. Finally, during the current process of updating the 2009 consensus statement, the influential Slick et al. (1999) diagnostic criteria for malingered neurocognitive dysfunction were updated by the original authors (Sherman et al., 2020), providing additional considerations for consensus participants as they worked on the current consensus statement update project.

For all of these reasons, there is an obvious need to update the 2009 consensus statement, which is the goal of this paper. Specifically, we intend to reinforce beliefs from 2009 that remain worthy in 2021, as well as provide additional guidance and recommendations regarding current expert opinion that can continue to assist clinical neuropsychologists in documenting the validity of examinee presentations and, when possible, the nature of invalid presentations.

### *General approach and timeline for 2021 update*

Unlike the first consensus conference statement, the current paper was not a product of a face-to-face meeting of invited experts. Rather, the approach of the current update was to develop and produce a statement derived from a detailed exchange of information via telephonic and electronic media.

In the spring of 2019, four of the original five 2009 work group chairs agreed to participate. Initial discussion, on March 21, 2019, between the work group chairs identified the following at the outset of the project: four topical sections for the update, a process by which consensus could be attained without requiring a face-to-face meeting, and a list of experts for the key subgroup topics. On March 26, 2019, a formal AACN Position Paper Proposal was completed and submitted to the AACN Publications Committee, subsequently endorsed by that committee, and forwarded to the Board of Directors, who voted for project approval. As in 2009, invitations to participate were extended to neuropsychologists who are engaged in relevant scholarly activity (e.g. peer-reviewed research, authored or edited books, continuing education for other professionals) devoted to the science and practice of neuropsychological assessment of validity. The goals were to: identify current key definitions of terms relevant to validity assessment; learn what experts believe should be reaffirmed from the original consensus paper, as well as new consensus points; and incorporate the latest guidance on the use of performance validity and symptom validity measures, as well as current application of the term 'malingering.'

Four subgroup topics were identified: Definitions and Differential Diagnosis, Performance Validity Assessment, Self-Report of Psychological and Somatic Symptoms, and Research Design and Statistical Issues. These categories were very similar to the original 2009 paper, with the exception that the previous paper had separate sections

on Psychological Symptoms and Somatic Symptoms, which are currently combined into a single section of Self-Report of Psychological and Somatic Symptoms.

### **Consensus process for statement update**

The four sections of the statement were initially prepared by work groups dedicated to their particular topic, with first drafts written by core writers and then shared for review and editing with additional internal work group experts. Each section draft was then reviewed by the co-organizers, with additional subsequent revisions within the work group. The resulting sections were then integrated by the co-organizers, who added initial drafts of the introduction and summary sections. Iterations of the entire document were reviewed multiple times by all participants. Ultimately, all participants in this consensus update had the opportunity to review, edit, and approve the entire completed 2021 statement, ensuring consensus among all 20 participants regarding sound practice guidance related to validity assessment. The consensus process concluded with a unanimous vote to support the present statement on January 21, 2021.

### **Definitions and differential diagnosis**

Since the previous consensus conference statement in 2009 (Heilbrunner et al., 2009), there has been a burgeoning clinical and research base devoted to the study of neuropsychological validity assessment and malingering. Over the past decade, multiple terms have either changed or been updated to more appropriately reflect currently prevailing views and paradigm shifts in the field. In the previous Definitions and Differential Diagnosis section, efforts were made to clarify various descriptors and terms that had previously been applied to convey generally accepted concepts at that time. The intent was not prescriptive, but aspirational. In this section, we will first summarize key points from the previous Definitions and Differential Diagnosis section that remain pertinent and then clarify current terms in a manner that facilitates an understanding of relevant concepts for future research and clinical practice.

### **Key points reaffirmed from the 2009 consensus conference statement**

A distinction was made between the processes of “detection” of invalidity and the “diagnosis” of malingering.

#### **Detection**

There are two areas of interest, both related to detecting attempts on the part of an examinee to demonstrate disability: (a) feigned and/or exaggerated symptom complaints and (b) feigned and/or exaggerated diminished capability.

- Indicators of symptom exaggeration, referred to as *symptom validity tests*, are typically incorporated into psychological tests that elicit self-report of symptoms and complaints.
- Stand-alone and embedded *performance validity tests* (tests referred to in 2009 as symptom validity tests or effort tests) are measures used to identify diminished

capacity. Stand-alone performance validity tests are those that are specifically, and usually solely, developed to address validity status. Validity tests that are 'built into' or derived from neuropsychological ability tests not primarily developed for purposes of assessing validity are referred to as *embedded* measures. Both are often included as part of a comprehensive neuropsychological test battery.

- When considering neuropsychological test performance, concerns regarding validity have often historically been thought to relate to the consideration of whether an examinee is malingering. However, simply equating suboptimal performance or test invalidity with malingering is an oversimplification. This is a complex conceptual issue, for which the 2009 consensus committee proposed a descriptive schema.
- Specifically, the process of detecting malingering is one in which consideration is given to multiple dimensions of underlying behavior, which differentiates it from other entities, such as factitious disorder and somatoform disorder and its variants (e.g. conversion disorder, cogniform disorder).

### Diagnosis

Within the 2009 consensus conference statement, the term *diagnosis* was used in a manner consistent with its use in the relevant neuropsychological literature at that time (e.g. Bianchini et al., 2005; Slick et al., 1999). In the version of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR; American Psychiatric Association, 2000) in use at that time, malingering was assigned a *V-code*, indicating that it is not a psychiatric illness and that a disease process is not implied with the designation. We agreed with the DSM-IV-TR that malingering is *not* a mental illness or disorder. Rather, malingering is a descriptive term that is reserved for instances of deliberate or intentional fabrication of symptoms or impairments in the interest of securing some form of external incentive(s). A classification of 'malingering' does *not* address the presence or absence of genuine symptoms or impairments, or related diagnosable conditions. Use of the term malingering is a determination, not truly a 'diagnosis' of a condition.

- In considering a designation of malingering, the clinician is explicitly making a determination of intent; more specifically, a determination of intentionally diminished capability and/or intentionally exaggerated symptoms with the goal of obtaining an external reward or benefit (e.g. financial compensation; avoidance of work or military service).
- The 2009 panel reached consensus that, through application of relevant psychological and neuropsychological science, clinicians can identify malingering in some examinees. To do this, the overall presentation of the examinee, including results of specific validity tests employed for the purposes of detection, should be considered.

For the identification of malingering, there are published classification systems (e.g. Bianchini et al., 2005; Slick et al., 1999) that represent the prevailing neuropsychological knowledge base. Such systems offer a reliable means of operationalizing decisions related to the determination of malingering. They are consistent with appropriate clinical neuropsychology practice guidelines (*AACN Practice Guidelines for*

*Neuropsychological Assessment and Consultation*; American Academy of Clinical Neuropsychology (AACN), 2007) and forensic psychology practice guidelines (*Specialty Guidelines for Forensic Psychology*; American Psychiatric Association, 2013), in explicitly recommending that examiners incorporate multiple sources of data and information when making determinations regarding malingering. Empirically-based assessment systems are recommended when making such determinations, as these provide increased reliability of the classification accuracy of findings of the various validity tests. It is also recommended that examiners be familiar with the psychological and neuropsychological literature related to the classification accuracy of validity tests and how well the sample of a given study generalizes to the individual being examined.

- Because a determination of malingering involves an explicit consideration of the *purpose* of a given behavior, the 2009 committee recognized that an important part of the assessment process involved a consideration of the *context* of the evaluation. In a routine clinical context, the *primary gain* of a patient is often relief of some form of physical or emotional symptom. In contrast, in forensic and some military contexts, the examinee typically has an external incentive to gain (i.e. secondary gain) by demonstrating impairment. To reach a determination that intentional exaggeration is produced as a manifestation of malingering versus some other condition (e.g. somatoform disorder), the clinician must determine that the examinee is attempting to appear impaired to obtain a secondary gain, recognizing that this is more likely to occur in forensic than clinical settings.
- The differential decision-making with regard to malingering is a process that: a) requires proactive actions on the part of the examiner (e.g. a priori selection of multiple PVTs and/or SVTs in the test battery), b) is based on objective criteria, and c) incorporates measures that have established classification accuracy. It is critical to combine professional judgment with the results of scientifically-validated validity measures in this process.
- The determination that malingering is present often involves the application of scientific results to a forensic question. This information can be used to assist the trier-of-fact (e.g. judge, jury) in a legal decision-making process. Neuropsychologists should remain mindful of the important difference between scientifically-based assessment decisions and legal adjudication. Moreover, it is important to recognize and respect the laws and local rules of the jurisdiction in which the services are performed when describing the behavioral presentation at issue. Finally, a competent determination of malingering requires that clinicians take into consideration the social, cultural, and legal contexts in which the behavior occurs (cf. Bush et al., 2020; Fujii, 2017).

### **Updates since 2009**

The professional literature on neuropsychological validity assessment has grown substantially since 2009, as documented by a PsycInfo search conducted on December 9, 2020. Searching this database (which includes book, journal, and dissertation

contents), for documents published between January 1, 2010 and December 9, 2020, and after removing items that were clearly false-positive findings (e.g. studies related to effort and reward systems), there were 432 documents containing the terms “effort” and “neuropsychological,” 860 documents containing the terms “malingering” and “neuropsychological,” and 408 documents containing the terms “performance validity” and “neuropsychological.” As discussed later in this consensus statement, the use of specific terms relevant to neuropsychological validity assessment has evolved over time. In 2012, the combination of “effort” and “neuropsychological” peaked at 62, whereas “performance validity” and “neuropsychological” was mentioned in only 7 published documents. By 2020, the combination of “effort” and “neuropsychological” had declined to 29, whereas “performance validity” and “neuropsychological” had increased steadily across the intervening years to being mentioned in 67 published documents. The combined search terms “malingering” and “neuropsychological” peaked in 2012 at 125, declining over time to a relatively stable frequency in recent years, which in 2020 was 57. Implications of these terminology preferences in the relevant professional literature are discussed at various points in this consensus statement.

A new version of the DSM (i.e. DSM-5; American Psychiatric Association, 2013) was released, which updated several diagnostic terms that have importance for validity assessment. Moreover, in 2020, Sherman, Slick, and Iverson published multidimensional malingering criteria for neuropsychological assessment, building upon their 1999 malingered neuropsychological dysfunction criteria (this is discussed in greater detail in the relevant sections that follow). Based on the additional literature that has been published, updates to the Definitions and Differential Diagnosis section were deemed warranted.

At the time of the 2009 consensus conference statement, the term *malingering* was sometimes broadly used by practitioners to imply invalidity. Consequently, ongoing clarifications of the term were necessary. The DSM-5 indicates that malingering is “the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives such as avoiding military duty, avoiding work, obtaining financial compensation, evading criminal prosecution, or obtaining drugs” (p. 726). The Malingered Neurocognitive Dysfunction (MND) criteria (Slick et al., 1999), which continue to be among the most frequently used criteria for identifying malingering by clinical neuropsychologists (Martin, Schroeder, & Odland, 2015), define malingering as a “volitional exaggeration or fabrication of cognitive dysfunction for the purpose of obtaining substantial material gain, or avoiding or escaping formal duty or responsibility” (p. 552). Notably, Sherman et al. (2020) chose to retain the term malingering in their updated conceptual model because they perceived that alternative terms (e.g. noncredible neuropsychological dysfunction, disability exaggeration) were not sufficiently precise to describe the entity known as malingering. In their 2020 multidimensional criteria for neurocognitive, somatic, and psychiatric malingering, they define malingering as “the volitional feigning or exaggerating of neurocognitive, somatic, or psychiatric symptoms for the purpose of obtaining material gain and services or avoiding formal duty, responsibility, or undesirable outcome” (p. 739).

The above definitions indicate that malingering is more than simply exaggerating or feigning of dysfunction. The exaggeration or feigning of dysfunction *must be motivated by an external incentive*. Additionally, to make a determination of malingering, a neuropsychologist must infer that the external incentive is driving the feigning behavior. In some evaluation contexts (e.g. personal injury litigation), external incentives are well established and can be inferred with a high degree of likelihood, while in other instances they cannot be determined or are less likely to be discerned (e.g. clinical settings in which evaluation results may later be used for forensic purposes, unbeknownst to the clinician initially). Forensic settings increase the likelihood of malingering, given that external incentives are always present, whereas causes of invalidity in non-forensic settings are believed to be more commonly due to other factors, such as an oppositional attitude toward testing or a somatoform condition, especially when external incentives are ruled out (Martin et al., 2015; Schroeder et al., 2016). Thus, malingering should be considered whenever validity testing suggests invalid responding, even if sporadic, and external incentives are present. However, invalid test findings should not be directly equated to malingering, especially in non-forensic settings, as validity measures can be negatively affected by factors that are not necessarily intentional or motivated by an external incentive. A determination of malingering most often requires a presentation in the context of an external incentive, pursuit of that external incentive, evidence of noncredible performance and/or feigned or exaggerated symptomatology, and evidence-based informed judgment by the practitioner, who should incorporate multiple pieces of information, including base rate data.

Like the term 'malingering,' the term 'effort' was often broadly applied at the time of the 2009 consensus conference statement. As one example of this, neuropsychological validity tests were previously referred to as 'effort tests' which can at times be misleading in that the results of validity tests are not synonymous with the common concept of effort. Effort, by common definition, typically refers to physical or mental exertion toward effective and successful completion of a task, not opposition to effectiveness on a task or task completion. However, an individual can effortfully resist being effective on a task and can expend effort in subverting or undermining the task, which runs counter to the intended meaning of the original labeling of validity tests as 'effort tests.' Also, effort exists on a continuum, ranging from maximum effort (100% effort), down through less than maximum effort (but more than no effort), to no (0%) effort. Neuropsychological validity tests are typically not constructed to measure a continuum of effort. Instead, they are constructed to identify *atypical patterns of performance* as compared to those of examinees who have genuine and significant cognitive dysfunction and produce valid test results (Larrabee, 2012). This is often accomplished by creating cutoff scores that are set to maintain specificity at 90% or better when applied to those individuals with evidence of significant cognitive dysfunction. A dichotomous validity test finding (i.e. scoring above or below a cutoff) is inconsistent with the construct of effort, which is on a continuum. Instead, the finding is reflective of a dichotomous decision as to whether the performance was so atypical as to invalidate performances, which cannot be viewed as an accurate reflection of examinee abilities.

Like the term 'effort test,' terms such as 'poor effort' to describe invalid test results can be misleading (Boone, 2013). Such terminology implies that an examinee simply did not put forth enough effort to produce valid results. Although this might be true in some cases of invalidity, it is not true in all cases. Particularly when malingering occurs, an examinee is rarely simply approaching testing with a lackadaisical attitude or low effort. Instead, the invalidity is commonly due to an intentional attempt to appear disabled, which, as noted above, can be an effortful process. A prime example of this would be within the context of significantly below chance forced-choice test performance. Such invalid test results are not reflective of poor or inadequate effort; instead, they are reflective of relatively high effort applied in a negative or counter-productive direction. Thus, to broadly refer to invalid results as being reflective of 'poor effort' would not be appropriate in such an instance. Additionally, in forensic settings, referring to invalidity in this manner could be erroneous, as implausible performances and evidence of probable malingering could be dismissed as simply being due to a 'lack of effort,' which discounts examinee accountability for the outcome.

Likely for the reasons stated above, use of terms such as 'effort test' and 'malingering test' have declined since publication of the 2009 consensus conference statement. Indeed, recent survey data indicate that the term 'effort measure' is preferred by only 14% of North American neuropsychologists and 8% of neuropsychological validity testing experts (Martin et al., 2015; Schroeder et al., 2016). The term 'malingering test' is preferred by only 2% of North American neuropsychologists and 0% of neuropsychological validity testing experts. Instead, more objective terms, such as 'validity test,' and 'performance validity testing' are preferred, the latter term by 87% of neuropsychology validity testing experts (Schroeder et al., 2016).

Since publication of the 2009 consensus conference statement, additional updates to relevant terminology have occurred. The term 'symptom validity test' (SVT) was previously used to refer to all neuropsychological validity indicators (both performance-based validity indicators and symptom report-based validity indicators). Larrabee (2012), however, indicated that using the term SVT to refer to all neuropsychological validity indicators is imprecise. He proposed that SVT be reserved for tests that assess the validity of reported symptom complaints (such as the Structured Inventory of Malingered Symptomatology [SIMS; Merckelbach & Smith, 2003] or validity scales within various inventories, such as the multiple forms of the Minnesota Multiphasic Personality Inventory e.g. MMPI-3; Whitman et al., 2020; Tylicki et al., 2020). Larrabee also proposed that performance validity test (PVT) be used to identify tests that examine the validity of stand-alone performance-based activities (such as the commonly used Test of Memory Malingering, Tombaugh, 1996, reviewed in Martin et al., 2020) and embedded validity measures, such as the commonly used Reliable Digit Span (RDS; Greiffenstein et al., 1994) and other Digit Span score derivations (reviewed by Webber & Soble, 2018). This terminology was quickly adopted; only two years later, nearly 75% of North American neuropsychologists and nearly 90% of validity assessment experts reported preferring using the term PVT to specifically refer to performance-based neuropsychological validity tests (Martin et al., 2015; Schroeder et al., 2016).

Finally, it is important to explicitly state that conditions other than malingering can be accompanied by distorted symptom report and presentation, particularly in non-forensic settings, in which external incentives are often less common than in forensic settings (Martin et al., 2015; Martin & Schroeder, 2020). Although some of the conditions that can cause distorted symptom presentation have been proposed outside of the DSM (e.g. cogniform disorder [Delis & Wetter, 2007], neurocognitive hypochondriasis [Boone, 2009a], stereotype threat [reviewed by Nguyen & Ryan, 2008]) and diagnosis threat [Suhr & Gunstad, 2002; Vanderploeg et al., 2014]), there are additional conditions included within the DSM.

At the time of the 2009 consensus conference statement, when the DSM-IV-TR was in use, conditions that could cause distorted symptom report and presentation due to somatization were largely listed under the “Somatoform Disorders” classification. In 2013, when the DSM-5 was published, the “Somatoform Disorders” classification was replaced by the newly termed “Somatic Symptom and Related Disorders” classification (American Psychiatric Association, 2013). Some of the clinical conditions under this classification are the same as those under the “Somatoform Disorders” classification, but others are different. All conditions under this new classification are said to share the common feature of having a prominent focus on somatic symptoms with associated distress and impairment related to the perceived symptoms. The DSM-5 notes that multiple factors might contribute to these somatic symptom conditions, including experiencing trauma in early life, having a major life stressor preceding the symptom onset, obtaining attention from having an illness, and having increased sensitivity to pain. It is also noted that some social or cultural norms might lead an individual to express emotional distress through physical symptoms (Maffini & Wong, 2014).

Regarding Somatic Symptom Disorder, the DSM-5 notes that this condition “offers a more clinically useful method of characterizing individuals who may have been considered in the past for a diagnosis of somatization disorder” (p. 310). Essentially, this condition is considered present when one or more distressing somatic symptoms interfere with daily living due to having: (a) disproportionate and persisting thoughts regarding the seriousness of symptoms, (b) persistently high anxiety regarding symptoms, and/or (c) excessive energy devoted to focusing upon symptoms. Somatic symptom disorder does not require an absence of a medical explanation for somatic symptoms; instead, this new diagnosis focuses upon atypical interpretation and/or presentation of somatic symptoms.

Other conditions under the “Somatic Symptom and Related Disorders” classification include conversion disorder (also referred to as functional neurological symptom disorder) and factitious disorder, which were both included in the DSM-IV-TR. Finally, there are also diagnoses of illness anxiety disorder (essentially, preoccupation with having or acquiring an illness), psychological factors affecting other medical conditions (psychological factors are thought to influence a medical condition or interfere with treatment of the medical condition), other specified somatic symptom and related disorder, and unspecified somatic symptom and related disorder. It is important to be aware of these clinical conditions, as well as other conditions that can result in distorted symptom report and presentation, when interpreting the results of neuropsychological evaluations, including validity measures.

## Performance validity assessment

### *Conceptual and operational definitions*

In the original 2009 consensus conference statement (Heilbronner et al., 2009), the content of the PVT section addressed assessment of test validity related to what were characterized in that era as “ability issues.” Specifically, this section included neuropsychological testing that involved assessment of domains such as attention, memory, language, processing speed, and sensorimotor function because these domains involved measurement of specific examinee abilities. Fundamentally, the tests associated with these domains all involve demonstration of examinee *performance* capacity, and, hence, are subject to consideration of performance validity. The relevant validity measures in these domains of ability are currently referred to as performance validity tests or PVTs. Performance-based tests are generally differentiated from self-report tests or inventories that involve direct examinee endorsement of symptoms (e.g. symptom validity tests or SVTs). However, self-report measures may include personal estimations of current or past capacities, or solicit performance ratings of the examinee from the perspective of a knowledgeable family member, significant other, or friend, in which case they are relevant to the content of this section of the paper. Consistent with self-report measures of non-performance abilities and symptoms, when self-report performance ratings are obtained, unless objective validity indicators are contained within the measures, the performance ratings cannot be assumed to be valid.

### *Key points reaffirmed from the 2009 consensus conference statement conceptual and operational definitions pertaining to assessment of performance capacity*

- Misrepresentation of performance capacities in any neuropsychological domain (motor, memory, etc.) represents response bias.
- When potential for external gain is present, and the valence of the response bias is negative (i.e. in the direction that would increase the likelihood or magnitude of external gain), malingering should be considered in the differential diagnosis.
- Scores beyond cutoffs, whether on stand-alone tasks or embedded indicators derived from ability tests, most commonly represent negative response bias. Marked inconsistencies between ability performances and abilities in everyday life could also represent response bias.
- Invalid neuropsychological test performance results: “(1) are not fully explained by brain dysfunction, (2) are not reasonably attributable to variables that may in some instances moderate (e.g. education, age) or may in some instances confound (e.g. fatigue, psychological conditions) performances on ability tests, and (3) are significantly worse than, or at least different in degree or pattern from, performance known to reflect genuine brain-based disturbances in neuropsychological abilities” (Heilbronner et al. 2009, p.1100).
- As part of addressing negative response bias and the possibility of malingering, review of records (including previous test results, if available), clinical interview,

and comparison of test results to behavior in the real world can be critical. Regarding interviewing, false and/or incomplete history is considered to represent more than normal error of omission or inexact provision of history, when related to the specific focus of a forensic consultation. Such inaccuracy on the part of the examinee could represent additional evidence of the negative influence of a context of secondary gain.

- As neuropsychologists judge validity of test data, they must do so by relying on tests and psychometric procedures that have proven validity. These tests and procedures are most often associated with multiple peer-reviewed scientific articles. The neuropsychological literature related to assessment of performance validity is well developed and expands substantially with new high-quality peer-reviewed research every year (cf. Sweet & Guidotti Breting, 2013; Martin et al., 2015). Practitioners are expected to maintain familiarity with the latest scientific literature in this area, assigning weight according to the rigor of the studies and the relevance of samples whose demographic and injury/illness severity characteristics most closely match the practitioner's case at hand. Moreover, it would be inappropriate to rely solely on PVT manuals without attention to newer cut-offs and interpretation strategies based on updated research.

### *Methods of evaluating performance validity*

- Use of psychometric indicators, rather than exclusive reliance on clinical intuition or qualitative observation of an examinee's approach to the task, is the most valid method for assessing the validity of a neuropsychological presentation.
- Stand-alone PVTs have been developed specifically to evaluate performance validity. The additional testing time required is warranted, considered *medically necessary* for clinical evaluations (Bush et al., 2005), and proven to be particularly useful within forensic evaluations, which have been shown to be associated with a high risk of invalid responding.
- Forced-choice stand-alone PVTs limit the examinee responding to one of a fixed number of responses. Although PVT results do not have to be significantly below chance to raise concerns regarding performance validity, results that are significantly below chance on forced-choice PVT measures indicate a willful avoidance of the correct response and support a conclusion of malingering when occurring within a secondary gain context.
- Non-forced-choice PVTs have been used effectively to assess a range of responses, such as random, unrealistically slow, implausibly inaccurate, and inconsistent patterns of responding that differ meaningfully from performances of individuals with well-documented disorders.
- Embedded performance validity indicators are measures that are contained within standard clinical ability tests and that have shown value in identifying noncredible or disingenuous performances.
- "As the number and extent of findings consistent with the absence or presence of response bias increases, confidence in conclusions regarding the validity of the examination is strengthened accordingly" (Heilbronner et al., 2009, p. 1107).

- In keeping with the fundamental belief that all test performances must be valid in order to be interpretable, neuropsychologists encourage examinees to perform to the best of their abilities.
- The validity of self-reported limitations of ability and symptoms of performance incapacity cannot be assumed and must be evaluated, particularly when the examination occurs in a forensic context or other context in which external incentives are present. When evaluating the validity of such self-reported limitations, clinicians include measures that possess an internal means of assessing response bias. Invalid responding and malingering pertaining to psychological conditions can occur independently of invalid responding and malingering pertaining to performance capacity, such as cognitive abilities (e.g. Nelson et al., 2007; Sherman et al., 2020). When psychological disorder and decreased performance capacity claims are both present, use of separate relevant assessment methods is required to ensure overall validity. These issues are discussed more extensively in the self-report section of validity testing with psychological and somatic disorders.

### *Malingering as applied to assessment of performance*

- Malingering is a real phenomenon; the term will, at times, be appropriate to use when examinations occur in a secondary gain context. Findings that are significantly below chance or involve *compelling inconsistency* (Bianchini et al., 2005) reflect deliberate misrepresentation of performance capacity. When present, the improbability of combined events can also reflect intent (Larrabee et al., 2007).
- Pertaining to performance validity, comparisons of test performances can be made with regard to: gross disparity with real-world behaviors, gross inconsistency with type or severity of injury, gross differences in behavior when aware of observation versus not aware, and gross inconsistency across serial testing that defies explanation as a genuine neurological or psychiatric condition. When malingered performance is present, it may or may not explain the overall presentation of the examinee and does not presumptively explain all prior or subsequent examinee behavior.
- Determination that performance is noncredible, invalid, or implausible can be made reliably when there is evidence of negative response bias. Performances that are not valid cannot be used as bases for opinions regarding attribution, nature and extent of deficiencies and disability, and/or treatment considerations.

### *Importance of evaluating performance validity when assessing performance capacities*

- Because specificity of PVT measures is set to reduce false positive errors, a positive finding probabilistically *rules in* invalid responding, but a negative finding may not *rule out* invalid responding (cf. Chafetz, 2021). However, a single invalid PVT performance within a large test battery could indicate a transient validity problem (Lippa, 2018) that might not support a general conclusion of malingering, which is a ‘big picture’ decision based more broadly on all available information, including test and non-test data. Moreover, as this decision is based on information beyond

just test findings, a determination of malingering may be appropriate even when all PVTs/SVTs are passed, depending on the entire test data set, the fact pattern of events/injuries in dispute, as well as broader historical information.

- Accurate results and interpretations require valid responding on every performance measure, requiring use of multiple validity measures throughout an examination (Boone, 2009b). The presence of invalid performance at any point makes a case for a conservative interpretive position that even valid-appearing performances may under-represent actual performance capacities.

### **Documentation of performance validity assessment**

- In their reports, neuropsychologists list the PVTs and validity assessment procedures that are utilized in evaluations. Clinicians explain the bases of their opinions to the extent required by the context, while avoiding inclusion of specific information pertaining to these measures that could preclude or undermine the validity of their future use.
- Whether or not a determination of malingering is made, terms such as *noncredible*, *invalid*, and *implausible* may be justified and can be used reliably on the basis of test results interpreted in the context of a specific case. When such terms apply, related performance data cannot provide a sound basis for: “(1) opinions with regard to attribution to the cause at issue (e.g. accident, injury), (2) the nature and extent of possible deficits and disability, and (3) guiding treatment or evaluating treatment effectiveness.” (Heilbronner et al., 2009, p. 1104)

### **Consideration of evaluation context**

- “The base rate of negative response bias varies as a function of setting. Adults presenting as litigants, defendants, or claimants in a criminal, civil, or disability proceeding or otherwise with motive to appear symptomatic (e.g. academic accommodations, drug seeking, excusing from military duty) show an increased risk (e.g. Ardolf et al., 2007; Chafetz, 2008; Greve et al., 2009; Mittenberg et al., 2002). For this reason, individuals seen in a forensic context should be administered measures that will assist in identifying or ruling out response bias (cf. Bush et al., 2005)” (Heilbronner et al., 2009, p. 1105).
- Routine clinical evaluations with adults and children generally have a lower risk of invalid responding, but the risk is not zero (Brooks et al., 2016; Kirkwood, 2015a; Rohling et al., 2015; Martin & Schroeder, 2020). The need to ensure valid responding applies to *all* cases.
- When the nature of claimed injuries of a patient seen in a routine clinical context could make that patient reasonably likely to become a litigant or claimant, the clinician should consider constructing an evaluation that is appropriate for the increased risk of response bias.
- Research has clearly shown that, absent reliance on PVTs, *clinical judgment* as a basis for identifying invalid performance is often inaccurate (Faust et al., 1988; Heaton et al., 1978). As such, if neuropsychologists do not use validity tests, they

are expected to document reasons for not doing so, as well as documenting the resulting limitations in validly interpreting findings.

- The presence of cultural, ethnic, and/or language factors known to affect performance results necessitates that clinicians adjust thresholds for identifying response bias (Nijdam-Jones & Rosenfeld, 2017; Salazar et al., 2021).

### ***New consensus considerations***

An extensive literature (published before and after 2009) attests to the continued effectiveness of stand-alone and embedded PVTs in discriminating credible from non-credible presentations. This empirical research base has played a significant role in shaping the beliefs and practices of clinical neuropsychologists over time. For example, in the years following the 2009 consensus conference statement, the two largest practice surveys of clinical neuropsychologists to date show tremendous acceptance of PVT use and a growing conviction in the strength of belief in PVT use. Specifically, with 1684 respondents answering relevant items pertaining to PVTs, Sweet, Benson, Nelson, and Moberg (2015) found that 91.2% of postdoctoral trainees, 88.5% of practitioners with no forensic experience, and 95.1% of practitioners with at least some forensic experience either “agree” or “strongly agree” that there is sufficient empirical research and knowledge to support the use of PVTs in practice. In a similarly-sized sample of 1699 respondents five years later (data collected in 2020), “agree” and “strongly agree” endorsements among these same three groups were 93%, 93.2%, and 95.3%, respectively (Sweet et al., 2021b). Interestingly, in the category of “strongly agree,” non-forensic practitioners increased from 45.9% to 62.2% and forensic practitioners increased from 64.5% to 75% relative to 2015. Only a single individual in the 2020 sample (1/1699) strongly disagreed with the notion that there is sufficient empirical research/knowledge to support PVT use in practice. Regarding actual use of PVTs, when asked in combination with SVT use, the 7.3% of practitioners who in 2015 did not use either PVTs or SVTs in *clinical* practice decreased to 2.8%. Related to *forensic* practice, approximately only 1% of practitioners in 2015 and in 2020 did not use either PVTs or SVTs. In this regard, beliefs are generally comparable between those who engage in forensic practices and those who do not, with respondents who are engaged in forensic practice more often reporting that they rely on *multiple* stand-alone PVTs and *multiple* embedded PVTs in clinical and forensic cases (Sweet et al., 2021b; Schroeder et al., 2016).

The large practice survey findings in 2015 and 2020 are generally congruent with other recent surveys that have examined PVT beliefs and practices among clinical neuropsychologists who work with both children and adults (Brooks, Ploetz, & Kirkwood, 2016; Martin et al., 2015), and forensic neuropsychology experts (Schroeder et al., 2016). Considering all of these relevant survey results in aggregate, it is clear that, among postdoctoral trainees and specialty practitioners, there is a *near universal acceptance* of the need to use stand-alone performance validity measures *and* embedded validity indicators in clinical *and* forensic evaluations. Related, an increasing majority of PVT research studies are now being carried out with clinical samples, not with mostly forensic samples as had been the case in the past (Suchy, 2019).

As discussed in the original consensus conference statement, clinicians are expected to encourage examinees to perform to the best of their abilities on neuropsychological tasks, and, in fact, recent survey data show that the vast majority of clinical neuropsychologists “often” or “always” encourage examinees to “try their best” (Brooks et al., 2016; Martin et al., 2015; Schroeder et al., 2016). Such encouragement is expected to be consistent with and not deviate from standardized test instructions. Practices that depart from standardized instructions can result in more sophisticated feigning, thereby impacting the effectiveness of PVTs. PVT classification statistics will not apply in contexts in which examinees are provided with information that was not provided to participants during validation studies; PVT effectiveness requires that examinees are unaware of their purposes and outcomes. This is generally true of all neuropsychological and psychological tests - examinees should be essentially naïve to specific tests, with the obvious exception of serial assessments. In this vein, examiners can consider a general admonition to examinees that lack of candor and accuracy in self-report or not demonstrating best abilities may negatively affect the examinee’s claim.

Verification as to whether examinees are performing to true ability in all neurocognitive skill areas requires use of multiple measures that cover various neurocognitive domains. In adult practice, well-validated PVTs are available that assess intellectual function, memory, processing speed, attention/working memory, visual perceptual/spatial skills, language, executive skills, and sensorimotor function. Such measures can be useful because invalid performance can take many forms; some individuals may feign memory impairment, whereas others may feign impairment in processing speed or attention. Rapidly evolving PVT research has been validating performance validity cut-offs for many standard neuropsychological instruments. Clinicians are expected to be knowledgeable regarding the validity indicators available for the neuropsychological tests they employ, including whether the indicators are relevant, as well as how they are to be applied, to the clinical disorder under consideration. It is preferable that clinicians attempt to select PVTs with the highest sensitivity to invalid test performance, while maintaining acceptable specificity, which is commonly set at 90%.

Single PVT failures can be found in credible patients and, with the exception of performance significantly below chance on a forced choice measure, should not, in isolation, be used to conclude that malingering is present (Victor, Boone, Serpa, Buehler, & Ziegler, 2009). Conversely, multiple PVT failures are not observed in credible populations, even those with cognitive impairment (Critchfield et al., 2019), with the exception of examinees who have well-documented severe functional disability (e.g. dementia) (e.g. Davis & Millis, 2014). Even in dementia patients and patients taking antipsychotic medications (cf. Ruiz et al., 2020), thresholds for identification can be modified to maintain adequate specificity, such that multiple PVT failures provide increasing evidence of performance invalidity (Martin et al., 2020). Empirical research currently does not support interpretation of multiple PVT failures as due to depression (Guilmette et al., 1994; Green & Merten, 2013), anxiety (Marshall & Schroeder, 2021), pain (Gervais et al., 2004; Greve et al., 2018), fatigue (Dorociak et al., 2018; Kalfon et al., 2016), PTSD (Demakis et al., 2008), medication effects (e.g. Schroeder & Martin, 2021a), or other putative conditions that might inaccurately be used to minimize or explain away PVT findings (Green & Merten, 2013; Schroeder & Martin, 2021a).

Normal, or even high scores, on some standard neurocognitive tests within a battery do not rule out the presence of malingering, as a sophisticated noncredible test taker may selectively choose tests on which to underperform within a neurocognitive battery, rather than electing to perform poorly on all tests (Boone, 2009b). It is also well recognized that an invalid presentation can be observed even when tangible external incentives may not be readily identified.

Given that PVT effectiveness is contingent on the naiveté of test takers regarding these tests, clinicians should be vigilant regarding protection of performance validity indicators. In particular, information such as test instructions and stimuli, scoring methods, and interpretation strategies and algorithms are to be withheld from non-psychologists.

As noted previously, there continues to be consensus that malingering is a valid concept for which there are identified criteria. Noting that applicable law and rules affecting limits of expert testimony from neuropsychologists can vary from one jurisdiction to another (cf. Richards & Tussey, 2013), in some practice locales and settings, practitioners may be advised to use alternative terms in expressing their diagnostic conclusions regarding the topic of malingering. That is, U.S. jurisdictions differ with regard to laws that may defer conclusions regarding the topic of malingering and related terms to the trier-of-fact, rather than to expert opinion (cf. Kaufmann, 2012). As local law and related courtroom procedures evolve over time, practitioners are advised to keep abreast of current practice expectations and rules of admissibility that determine limits of opinion witnesses, and, as appropriate, may need to convey results descriptively, relying on more general terms, such as “invalid” and “noncredible.” It is also not appropriate to label a claimant as having committed “fraud” on the basis of PVT findings, as a finding of fraud is a legal determination not made by a practitioner.

It is well recognized that invalidity can be observed, albeit less often, even when tangible external incentives may not be readily identified. Although PVTs were originally developed and researched with a focus on forensic contexts, there is now a more explicit and growing appreciation that PVTs are needed in assessments in routine clinical contexts as well (Schroeder & Martin, 2021b). Clearly encouraged in the original 2009 consensus conference statement, there is currently even more reason to believe that performance validity testing is needed in all cases, whether clinical or forensic. For example, Martin and Schroeder (2020) surveyed practicing neuropsychologists, the majority of whom were board certified, regarding invalid responding in clinical non-forensic cases. Survey participants reported a general estimate of an invalid response rate of 15%. However, with more specific questioning, subsamples of clinical non-forensic cases were estimated to range from invalid response rates of 5% to 50%, depending on case-specific circumstances. Notably higher estimates were noted for patients with somatic symptom disorder, conversion disorder, and “medically unexplained symptoms,” as well as patients not currently involved in pursuit of external incentives (i.e. not presently being evaluated in a secondary gain context), but who had the potential for, or were considering, such pursuit. In this regard, the NAN position paper on the ‘medical necessity’ of validity assessment (Bush et al., 2005) and AACN Practice Guidelines (2007) continue to offer sound guidance supporting the evaluation of validity in all neuropsychological evaluations. And, without question, in

2020, postdoctoral trainees and practitioners in clinical neuropsychology are adhering to this guidance (Sweet et al., 2021a, 2021b).

In the 2009 consensus conference statement, the criteria for malingered neurocognitive dysfunction authored by Slick et al. (1999) were viewed as influential and a positive step toward improving diagnostic agreement across clinicians. That viewpoint has not changed. Recently, the so-called 'Slick' criteria have been updated by the original authors (Sherman et al., 2020), and retitled as criteria for malingered *neuropsychological* dysfunction, rather than *neurocognitive* dysfunction. In the intervening years, the original Slick criteria had been scrutinized and, in some sense, reconsidered by a number of expert authors, including the original authors, often related to forensic application (cf. Boone, 2007; Larrabee et al., 2007; Slick & Sherman, 2012, 2013). There has been general agreement that the original Slick criteria needed modification, related to a number of issues, stated by the authors as including: being focused on effort, rather than compliance; being overly focused on cognitive malingering, as opposed to psychological and somatic malingering, which are also addressed by neuropsychologists; not giving enough consideration to self-report information, and too narrowly relying on forced-choice procedures as definitive evidence. Other sections of this consensus conference statement will address applications to non-performance-based testing.

Regarding performance-based validity assessment, there is consensus that the updated criteria from Sherman et al. (2020) offer guidance that is, in part, consistent with current conference participant views, including: (1) distinguishing between PVTs and SVTs; (2) broadening the consideration and scrutiny of self-report information in the context of other sources of real-world examinee information (e.g. medical records, video and social media) and expected disease course; (3) being in accord with the prior 2009 consensus conference statement support for the concept of 'compelling inconsistencies'; (4) expanding the range of external incentives from purely compensation-seeking; (5) acknowledging that malingering presentations can involve mixed symptoms (e.g. two or more of cognitive, somatic, or psychiatric presentations); (6) recognizing that both PVT and SVT results can shed light on the validity of cognitive tests; (7) acknowledging that simulation studies alone are not sufficient for the validation of PVTs/SVTs and that known-groups studies are necessary for their validation; (8) acknowledging the strength of evidence for PVTs/SVTs failed at cutoffs with 100% specificity; and (9) noting that surpassing thresholds on multiple PVTs can have an effect on neuropsychological performance that is indistinguishable from significantly below-chance performance on a single PVT.

Although in many ways compatible with consensus participant views, there is concern regarding a few points made in Sherman et al. (2020). For example, their new diagnostic criteria require evidence for "marked discrepancies" (e.g. disease course and symptoms not consistent with claimed condition, etc.) in addition to external motive to feign and failed PVTs and elevated SVT scores. Although such information is found in the large majority of cases, and is particularly compelling when present, it is possible that in some situations in which there is a paucity of records, it may not be possible to formally document "marked discrepancies." However, there is consensus that external incentive and multiple failed PVTs point strongly to a noncredible

presentation that may indicate malingering, if presence of a genuine marked functional disability can be ruled out.

### *Additional new consensus considerations for pediatric practice*

Attention to validity testing in pediatric populations pales in comparison to that in adults, which is why the 2009 consensus conference statement did not substantively address pediatric issues. However, the pediatric empirical literature on validity assessment and malingering has now grown large enough to support a recommendation that practitioners should use PVTs routinely in neuropsychological evaluations of school-aged children and adolescents. This recommendation is supported by the following points (Kirkwood, 2015a): (1) a sizable developmental psychology literature demonstrates that increasingly sophisticated deceptive behavior occurs throughout childhood and into the adolescent years; (2) as children become older, even professionals have difficulty subjectively identifying deceptive behavior, similar to what is apparent for adults; (3) both case reports and case series have documented that children can, and do, feign many types of physical, psychiatric, and cognitive difficulties in healthcare settings; (4) although case series indicate that noncredible responding does not commonly occur in most pediatric clinical neuropsychological settings, it does happen consistently; (5) certain pediatric populations have been found to be at elevated risk for noncredible responding (e.g. youths seen for persistent problems after concussion and for Social Security Disability determination evaluations); (6) a number of embedded and stand-alone PVTs are now well-validated for use in school-aged children (cf. Clark et al., 2020; Colbert et al., 2021; Kirk et al., 2020; Kirkwood, 2015b); and (7) despite the fact that well-established PVTs are insensitive to all but the most severe ability-based deficits, PVT failure in children, similar to what has been found in adult populations, accounts for substantial amounts of variance on performance-based tests.

Motivations underlying noncredible test performance vary. In adult populations, a common explanation is malingering when external incentives are readily identified. However, whether children malingering relates to their developmental capacity to understand the consequences of their behavior, as well as the array of incentives that they may be seeking. Many of the motivators in common descriptions of malingering were written with adults in mind (e.g. financial compensation, avoiding work, evading criminal prosecution) and can be expected to occur less frequently in pediatric populations. It is important to note, however, that noncredible performance still happens in child and adolescent evaluations. When it does, practitioners should consider external incentives that may be more relevant to youth populations in attempting to understand the behavior. Motivators such as avoiding schoolwork, securing academic accommodations, avoiding social stressors (e.g. bullying), or getting out of a sport commitment are just a few. When working with children and other vulnerable populations, “malingering by proxy” must also be considered as a possible explanation for noncredible presentations. Malingering by proxy occurs when an individual is coerced into noncredible behavior by a caregiver/parent for purposes of external gain. Factitious disorder, driven by the youth or a caregiver, may also help to explain some

noncredible behavior in pediatric settings, including the type of behavior associated with PVT failure (Chafetz et al., 2020).

Engagement with performance-based testing can wax and wane for various reasons (e.g. fatigue, attentional deficit), with natural fluctuations more common in certain examinees, including children. An important distinction is whether the performance is credible or noncredible, given the individual's developmental, neurologic, and psychiatric status. The use of well-established PVTs is a primary means of making this distinction, as natural fluctuations of variables that impact task engagement do not typically cause PVT failure.

### *Future research directions for PVT application*

By definition, malingering is an intentional act of deception regarding illness or impairment of function. The very nature of known (criterion) groups research permits the inference of malingering via the development of accuracy statistics for PVTs and SVTs. The known-groups format relies on the use of established guidelines (e.g. Sherman et al., 2020) for the identification of the malingering group in the development of a PVT/SVT. When the finding from a PVT/SVT is obtained in a new individual case, these error rates are then applied to update the probability of malingering in that individual case from the baseline probability of malingering. Thus, the known-groups research design, with its adherence to established guidance for the identification of the malingering group, permits an inference about this intentional act of deception.

Since the 2009 consensus conference statement, research into validation of embedded PVTs has been particularly prolific, and future research should continue to develop validated performance validity measures for virtually all standard neuropsychological measures so that performance validity can be evaluated in 'real time' for every test. Validation of increasing numbers of embedded PVTs results in an increasing amount of PVT data available within each neuropsychological evaluation. In general, as PVT failures increase in numbers within a single evaluation, there is increasing confidence regarding the presence of invalid responding. Research has shown that use of increasing numbers of PVTs raises the false positive rate only slightly (Davis & Millis, 2014; Larrabee et al., 2019), but additional research is needed to guide the clinician in how to interpret results of large numbers of PVTs. For example, preliminary research shows that if data from nine PVTs are employed, failure on one is expected in credible test takers (Davis & Millis, 2014); therefore, two or more failures might be consistent with invalid performance. Because there are still minimal data regarding the expected number of failures when more than nine PVTs are failed, we do not have consensus regarding specific interpretive implications related to the number of invalidated PVTs among the total number of PVTs administered, when a high number of PVTs is administered. In this regard, the current consensus panel believes that the specific ratio recommendation made by Sherman et al. (2020) is premature.

More research is needed to demonstrate the base rate of noncredible responding and PVT classification statistics across (a) child samples ranging more widely in ability level, (b) presenting condition, and (c) evaluative context (e.g. some clinical, forensic, and school settings). The base rate of noncredible findings in children brought by

caregivers/parents for Social Security disability evaluations has been examined (Chafetz, 2008). Guidelines for mandated reporting of child abuse/neglect due to malingering-by-proxy have been established (Chafetz & Dufrene, 2014).

Additional research into new PVTs for children and adolescents is in order, to establish a more complete armamentarium of pediatric-specific measures similar to that which is available for examination of adults. To ensure that repeated monitoring of performance validity can occur throughout evaluations of children and adolescents requires availability of multiple embedded PVTs. Test publishers and independent investigators are encouraged to develop post-release research on new pediatric tests to develop embedded PVTs across performance-based test domains.

Although a longstanding and important historical issue, it is with increased emphasis that we view test security as one of the most critical issues in performance validity testing at the present time. Future research should focus on development of performance validity assessment methods that are robust to test security violations. Tests administered by computer offer potential in this regard. For example, a computer-administered task may not be harmed by a court order that allows audio or video recording of an examination. Moreover, computer-administered PVTs need not produce formal *test data sheets*, in which case court orders to release this specific information would not result in divulging the level of test information that would cause harm to future test use, if released to non-psychologists. The general development of standard test data sheets that do not include test stimuli, instructions, scoring methods, and other information that would invalidate future test use if released to non-psychologists is also strongly encouraged.

Research is also needed to develop performance validity assessment methods that are particularly robust to coaching. Examples of such methods include the following: (1) combining scores across various tests to detect noncredible patterns of scores, via statistical approaches such as logistic regression, that incorporate both intact and low scores (i.e. patterns of intact and impaired scores can likely be identified in patients with actual brain dysfunction that are not mimicked by test takers who are feigning), (2) performance validity methods that employ automatic priming paradigms that are performed normally by patients with severe memory impairment, but are abnormal in test-takers feigning impairment because the types of responses found in patients with brain dysfunction are less obvious and intuitive, and (3) routine incorporation of response times in PVT paradigms, because time scores have been shown to add to test sensitivity (Kanser et al., 2019).

## Self-report of psychological and somatic symptoms

### *Key points reaffirmed from the 2009 consensus conference statement*

Psychological, somatic, and cognitive complaints are core diagnostic criteria for the disorders and conditions delineated in the Diagnostic and Statistical Manual of the American Psychiatric Association (DSM-5; American Psychiatric Association, 2013). This section of the document represents an update to the 2009 consensus conference statement (Heilbronner et al., 2009), which remains foundational in its content and

recommendations. The following relevant points from the original 2009 consensus statement are reaffirmed:

- Self-report of psychological and somatic symptoms may be biased, exaggerated, minimized, or false. This potential of misrepresentation is especially true in compensation-seeking assessment contexts. Genuine psychopathology/somatic symptoms may co-exist with exaggerated or feigned symptoms, and examiners should attempt to differentiate one from the other to the extent possible.
- Well-validated objective measures should be utilized for assessing self-reported somatic and psychological problems. The degree of the examiner's confidence regarding assessment conclusions depends on the absence or presence of response bias as demonstrated on multiple, validated assessment measures. Assessment instruments providing the most scientifically supported and current methodologies should be utilized.
- Cognitive, somatic, and emotional complaints are common in individuals undergoing forensic assessment, and those domains require validity assessment and analysis. Multiple symptom validity measures (e.g. multiple validity scales contained within a broad personality test or multiple symptom self-report tests that contain validity scales) should be utilized during the evaluation. When available and relevant, examiners are encouraged to use actuarial/base rate data in their analyses and determinations.
- Thorough consideration of the examinee's history and cultural and ethnic factors remains foundational in all determinations of genuine versus invalid self-reporting of cognitive, somatic, and psychological complaints.
- As noted in the 2009 consensus statement and similarly retained within the section of the current statement with regard to assessment of abilities, "It is the totality of the examinee's presentation that should be taken into account when assessing the validity of claims of psychopathology and disability/impairment" (Heilbronner et al., 2009, p. 1112).

### ***New consensus considerations***

Symptom and performance validity assessment is a critical component of the standard of care in forensic and clinical assessment contexts, as neuropsychologists recognize the need for confidence in the self-report data collected from examinees. Examinees presenting with histories and/or self-reports of psychopathology are common in both clinical and forensic neuropsychological assessment contexts. It is well established that subjectively-reported complaints of cognitive dysfunction may be associated with psychiatric symptoms (cf. Markova et al., 2017) and actual cognitive dysfunction may accompany or result from extant psychiatric illness (East-Richard et al., 2020). Research shows a relationship between performance validity assessment and symptom validity assessment, indicating that cognitive and psychiatric feigning are related (Bianchini et al., 2018; Gottfried & Glassmire, 2016). However, although there is some evidence that failure on PVTs is associated with exaggeration, as indicated by specific SVT scales (Larrabee et al., 2017), failure on one does not *necessarily* predict failure on the other

(Van Dyke et al., 2013). Moreover, scores on SVTs or PVTs may be domain specific. For example, an examinee may exaggerate or feign cognitive complaints while minimizing or denying psychological or somatic complaints.

There is a diversity of disorders seen in forensic and clinical settings. Forensic cases range from those with a high base rate (e.g. somatic symptoms, anxiety, and depression) to those that are somewhat rarer (e.g. schizophrenia, bipolar disorder), while in clinical contexts attention-deficit/hyperactivity disorder (ADHD) and learning disorders (LD) are common, especially in school-aged children and young adults. Just as in the assessment of neurocognitive performance, neuropsychologists endeavor to evaluate the accuracy (i.e. validity) of examinees' symptom reports with regard to the presence of a diagnosable condition, its severity, and potential effects on functional status and behavior. Subjectively reported psychiatric, somatic, and cognitive symptoms require objective validity analysis, and examiners are encouraged to be aware of possible symptom minimization, denial, magnification, or feigning.

### *Overview of somatic symptom disorders*

Somatoform disorders represent a category of psychopathology generally falling in the domain of 'unexplained medical illness' (cf. Carson et al., 2016; Scamvougeras & Howard, 2018). Such disorders are common, representing 20% to 30% of general medical and neurologic patient visits, respectively (Boone, 2017). As a category, these disorders share the commonality that a medical/physiological cause for the reported symptoms has not been determined or does not fully account for the condition (American Psychiatric Association, 2013). Relevant to the present consensus statement, validity testing has been found to make an important contribution in evaluations of individuals with medically difficult-to-explain and unexplained medical symptoms (e.g. Kemp et al., 2008; Lockhart & Satya-Murti, 2015)

There are numerous forms of somatic symptom disorders that are accompanied by symptoms such as excessive illness-related anxiety, pain, weakness/paralysis, speech symptoms, and sensory symptoms. For example, individuals with chronic pain commonly present in neuropsychological assessment contexts, often with comorbid psychological, cognitive, and/or other somatic symptoms (Greve et al., 2018; Sherman et al., 2020). Self-reported pain symptoms may be exaggerated and/or malingered (i.e. Malingered Pain Related Disability [MPRD] Bianchini et al., 2005), requiring thorough validity assessment. There is also ample research evidence that, in addition to over-reporting symptoms of pain, such examinees may also over-report or feign cognitive and/or psychological symptoms (Bianchini et al., 2005; Greve et al., 2013, 2018).

There are also unexplained 'disorders', such as toxic mold syndrome, multiple chemical sensitivity, sick building syndrome, chronic fatigue syndrome, and fibromyalgia, among others (Boone, 2009; Boone, 2011; Henry et al., 2018; McCaffrey & Yantz, 2007; Suhr & Spickard, 2007). Psychogenic non-epileptic seizure (PNES) disorder represents a particularly interesting somatoform-type illness, as it can occur in individuals with genuine seizure disorders (Locke et al., 2017; Williamson et al., 2015). Related, an important point of consideration is that a somatic symptom disorder can be comorbid with genuine medical and other mental disorders.

Post-concussion syndrome (PCS), a term that some providers apply when there is a persistent symptom cluster that follows after mild traumatic brain injury (mTBI), is a disorder that is accompanied by multiple non-specific somatic, emotional and cognitive features (Broshek et al., 2015). The etiology of PCS is most often determined to be multifactorial, and, when chronic, commonly believed to be primarily psychological (Belanger et al., 2018; Iverson, 2012; Ponsford et al., 2012; Wood et al., 2014). Notably, non-brain injured individuals have been found to report the same symptoms, with comparable frequency and severity (e.g. Kashluba et al., 2004; Mickevičiene et al., 2004; Lange et al., 2020). The sequelae of concussion on somatic, cognitive, and emotional functioning are considerably contentious in litigation contexts and tend to be primarily represented by two opposing points of view: those who view concussion as directly leading to long-term cognitive, emotional, and somatic symptoms versus those who view outcome from concussion as benign and long-term symptom report as due to other factors. Although considerable research indicates that the vast majority of mildly concussed individuals return to baseline, most often within the initial days and weeks after injury, with fewer requiring as much as a month or two after injury (Belanger et al., 2018; Dikmen et al., 2010; McCrea, 2008), a very small percentage of adults continue to report subjective lingering symptoms (Iverson, 2007; Lange et al., 2012; Ruff et al., 1996), even when objective neurocognitive testing is normal. Critical to the present discussion is the repeated empirical finding that litigation increases the likelihood for invalid symptom report (French et al., 2018; Lange et al., 2010; Vanderploeg et al., 2014).

Over-reporting of physical symptoms is a common feature of somatic symptom presentations in forensic contexts and other contexts in which external incentives are present. Invalid responding on SVTs among these individuals is common and is associated with exaggerated and noncredible subjective reports of somatic symptoms (Wygant et al., 2007; Wygant et al., 2009; Boone, 2017, 2018).

Factitious disorder may be conceptualized as somatoform-like in that in both disorders, individuals organize their lives around the assumption of a sick role identity and toward seeking medical care (Boone, 2013, 2017). The main difference is that individuals with factitious disorder knowingly feign illness, ostensibly for attention and even in the absence of an apparent external incentive, whereas somatoform patients believe they are actually ill (Krahn et al., 2008). Children have also been known to feign illness, present with factitious disorders, and be negatively influenced by coaching from a parent or other family member (via proxy) (e.g. Shaw et al., 2008). While survey research indicates that pediatric neuropsychologists tend to utilize SVTs in both forensic and clinical assessment contexts (Brooks et al., 2016), data suggest that use is lower for some child referral questions, such as ADHD (DeRight & Carone, 2015), and that there are limited data on the validity of pediatric SVTs for neuropsychological presentations (Kirkwood, 2015b).

Although some debate continues as to whether factitious disorders and frank malingering are distinct or really one-in-the-same (McCullumsmith & Ford, 2011), some professionals (and the DSM-5) take the position that the locus and type of incentive (external vs. internal) suggests that they are distinct. Others, however, point out that adoption of the sick role, a typical internal incentive, commonly brings with it *external*

incentives (e.g. avoidance of responsibility; Bass & Wade, 2019). Nevertheless, both factitious disorder and malingering share the common feature of volitional noncredible symptom reporting (i.e. *illness deception*) (Chafetz et al., 2020). Attempting to lend greater clarity to the diagnosis of factitious disorder than is found in the DSM, Chafetz and colleagues recently proposed diagnostic criteria and a heuristic for relevant research on this topic (Chafetz et al., 2020).

### *Overview of psychological disorders*

There are multiple psychological disorders that share anxiety as their common feature. Among these disorders are generalized anxiety, phobias, and panic attacks. Examinees may manifest any number of symptoms from the general domain of anxiety. Moreover, anxiety is also a frequent concomitant in other mental disorders and is encountered in many forms of cognitive impairment. Anxiety is often identified in neuropsychological assessment contexts, either as the main presentation, or comorbid with cognitive, medical, and/or mental health disorders. Although cognitive complaints are common in individuals with anxiety disorders, there is little evidence for actual cognitive impairment, despite these complaints (Mutchnick & Williams, 2012; Kizilbash et al., 2002; Gass & Curiel, 2011). The frequent discordance between subjective cognitive complaints in anxious patients and their objective findings is typically not a sufficient basis for questioning test response validity (i.e. test findings are often normal). In some cases, panic disorder, phobias, and obsessive-compulsive disorder may affect cognition (Airaksinen et al., 2005; Kuelz et al., 2004). In such instances, results of validity testing will be important in determining interpretation of findings.

One of the more prevalent disorders seen in compensation-seeking assessment contexts is posttraumatic stress disorder (PTSD). As with any mental disorder, information concerning its symptoms and presentation can be easily researched, providing fodder for examinees wishing to overreport or overtly feign symptoms. Research indicates that PTSD is particularly easy to feign (Ray, 2014; Matto et al., 2019). Nevertheless, symptom overreporting can be psychometrically identified, utilizing comprehensive validity analysis (Andrikopoulos, 2018; Andrikopoulos & Greiffenstein, 2012; Goodwin et al., 2013).

Depressive disorders are common and span a range from mild to debilitating. As with most mental health disorders, depression may present as a distinct diagnosable disorder (e.g. Major Depressive Disorder, Persistent Depressive Disorder), or may present comorbidly with other mental health disorders (e.g. substance abuse) or medical disorders such as chronic pain or chronic illness. Like PTSD and other mental disorders, depression is relatively easy to exaggerate and/or feign (Marion et al., 2011). Depression and depressive features have inconsistently been found to be associated with cognitive dysfunction (e.g. Murrough et al., 2011; Marcopulos, 2018), and may increase misdiagnosis of cognitive impairment (e.g. Edmonds et al., 2014), possibly because performance validity is not always examined in published research studies (Green & Merten, 2013). As with anxiety, the mere finding of a discordance between subjective complaints and objective findings, which may be normal, is not sufficient in identifying invalid responding. Instead, results of validity testing will provide better guidance for interpretation.

Psychotic disorders are somewhat less common in neuropsychological assessment contexts and are primarily seen when acute symptoms have resolved. Both bipolar disorder and schizophrenia may impair cognition (Cotrena et al., 2017; Marcopulos, 2018). There is little doubt that psychotic symptom reports can be exaggerated, although successfully feigning a psychotic disorder may be more difficult, but discernible by an experienced examiner (Morgan et al., 2009; Resnick & Knoll, 2018). Nevertheless, exaggerated claims of severe psychopathology and/or malingered psychosis are not uncommon in the criminal forensic setting (Denney, 2008). Use of specific validity assessment instruments and scales to identify symptom over-reporting in such assessments is indicated (Arbisi & Ben-Porath, 1995; Marion et al., 2011; Nelson et al., 2010; Rogers et al., 2010; Sharf et al., 2017).

Children, adolescents, and young adults may present with any of the previously described disorders, as well as disorders that are more common among this age group. Persons with known or suspected ADHD and LD are often referred to neuropsychologists for assessment. ADHD typically manifests cognitively as impaired attention and executive dysfunction, whereas with LD, academic skills and executive functions are often affected. Examinees with these possible conditions whose intent is to receive accommodations in educational settings (e.g. for standardized testing, such as SAT, GMAT, LSAT) or prescriptions for stimulant medications may exaggerate or feign cognitive impairment and psychological dysfunction to attain such goals (Harp et al., 2011; Harrison et al., 2021; Hurtubise et al., 2017), underscoring the importance of including SVTs and PVTs in these evaluations. Moreover, both LD and ADHD commonly have emotional concomitants, such as emotional dysregulation, anxiety, and depression (Wehmeier et al., 2010; Sobanski et al., 2010; Sjöwall et al., 2013), necessitating SVT assessment. Clinicians should duly note that relevant research has shown a high base rate of PVT and SVT failure in ADHD and LD assessment in older teenagers and young adults (Harrison & Armstrong, 2016; Harrison et al., 2015; Marshall et al., 2010; Williamson et al., 2014).

This overview represents merely a sample of the psychologically and somatically symptomatic presentation of examinees frequently encountered in forensic and clinical neuropsychological assessment contexts. It is recognized that some degree of overlap exists among these disorders, both conceptually and in terms of presentation, and comorbid conditions are common. It is also recognized that there are other mental health conditions that may be encountered in clinical and forensic examination contexts that are not specifically mentioned in this document (e.g. personality disorders). In all cases of psychological evaluation, the value of appropriate validity analysis of reported symptoms cannot be overemphasized.

### ***Assessment methodology***

Examinees presenting with cognitive, psychological, somatic, or behavioral concerns require comprehensive assessment of their reported symptoms in both clinical and forensic examinations. In addition to neuropsychological tests and PVTs (used to assess validity when cognitive symptoms are reported), comprehensive assessment often includes review of medical and academic records, clinical interview, completion of

brief or detailed self-report questionnaires that include embedded validity scales, and may include free-standing SVTs.

### **History and records**

Neuropsychologists should carefully review the medical and/or academic records of examinees, recognizing that some details and symptoms are subjectively reported and recorded in notes. Experience indicates that providers across disciplines, including physicians, in various clinical contexts generally accept patients' subjective reports at face value, documenting them in the record in an uncritical manner. Some examinees will have seen multiple providers, subjectively reporting symptoms to all. Commonly, providers form a diagnostic impression based on these subjective reports, sometimes with the use of self-report symptom questionnaires that lack validity indices. Most skilled neuropsychologists understand that uncritical acceptance of symptom reports is not acceptable, particularly in secondary gain contexts (Rosner & Scott, 2016; Heilbronner et al., 2009; Morgan & Sweet, 2008; Sweet et al., 2018; Kirkwood et al., 2011). In fact, there is a long history of research demonstrating that the accuracy of clinician 'impressions' of credible symptom reporting alone is poor and, sometimes, even no better than chance (Dandachi-FitzGerald, Merckelbach, & Ponds, 2017; Faust et al., 1988; Heaton et al., 1978). Similarly, diagnostic impressions of examinees based entirely on their self-reported medical or school record, without documented objective validity assessment, are likely to be erroneous in both children and adults (Kirkwood & Kirk, 2010; Martin et al., 2015). For example, an examinee might minimize/deny prior academic concerns when feigning current cognitive/academic problems secondary to a recent mTBI, or might exaggerate the severity of their mTBI relative to objective records from the time of the injury.

In some cases, well-meaning providers inadvertently reinforce false conceptions in patients, unwittingly promoting an iatrogenic "disorder." This is not unusual in mTBI cases, when emergency department staff may fail to provide sufficient information and reassurance regarding recovery expectations (Bender & Matusiewicz, 2013; Carone, 2018; Kirkwood & Kirk, 2010). This can also be seen in the psychoeducational setting, when students are informed that they have a neurodevelopmental disorder, such as ADHD or dyslexia, despite lack of evidence consistent with the diagnosis, and then receive accommodations (Suhr & Wei, 2017). In situations like these, both the medical record and examinee beliefs may be erroneous.

### **Prior assessment results**

When the medical or academic records contain neuropsychological/psychological assessment reports, examiners should compare previous test results to their own, including analysis of prior scores, and underlying *raw data*, when possible. If previous assessments utilized SVTs and/or PVTs, these should be carefully reviewed and compared against current results. When relevant records indicate worsening of reported symptoms within a context of apparent appropriate, evidence-based treatment, examiners should assess the psychosocial context to identify possible causes and credibility

of symptom exacerbation. Relevant research indicates that appropriate treatment is typically associated with symptom improvement in many mental disorders (Goldfried, 2013). Lack of assessment regarding the validity of self-report in previous evaluations should raise concerns about their interpretability.

### *Clinical interview*

Neuropsychological evaluations commonly include a comprehensive clinical interview of the examinee and interviews of collateral sources where possible and appropriate, with the understanding that the accuracy of the information gathered is unverified and should not be used to over-ride objective test data. The examiner should carefully document the reported symptoms, behaviors, and mental status of the examinee. Particular care should be taken to detail the emergence of symptoms, noting the psychosocial, environmental, cultural, and familial/genetic factors that may be operative in the development of psychopathology. Where symptoms are thought to credibly represent an extant mental disorder, examiners judge the severity and nature and the effect of such symptoms on functional capacities, behaviors, and quality of life.

Structured interview protocols may not include assessment of the validity of self-reported symptoms or impairment. Examiners should consider this when selecting and using structured interview protocols.

### *Administration of SVTs*

Examiners administer SVTs as part of a systematic process for validating reported symptoms. Free-standing SVTs are available, as well as self-report measures with embedded SVT scales. Research with adults indicates excellent ability to determine overreporting/exaggeration of symptoms and to discriminate credible from noncredible examinees with certain instruments that contain validity scales (Ben-Porath et al., 2009; Greiffenstein et al., 2007; Gervais, Ben-Porath, Wygant, & Sellbom, 2010; Ingram & Ternes, 2016; Sellbom & Bagby, 2010). The use of SVTs provides an objective means of detecting exaggerated/feigned psychological and somatic complaints (Ben-Porath & Tellegen, 2008; Boone, 2017; Larrabee, 2007). The use of multiple SVTs, including psychological tests that have multiple embedded validity indicators, provides the most scientific basis for opinions provided in secondary gain contexts. That is, multiple SVTs, or the use of multiple SVT data points, such as multiple internal over-report validity scales (e.g. MMPI-2-RF/MMPI-3), are recommended for forensic examinations (Sherman et al., 2020). Such guidance is consistent with the vast majority of current practices of neuropsychologists when conducting clinical *and* forensic evaluations (Sweet et al., 2021b).

As previously noted, examinees may, at times, minimize or deny a previous history of mental health treatment or medical disorders or minimize/deny their past and current psychological or medical symptoms. Motivation to attribute pre-existing symptoms and dysfunction to more recent/current events, especially when external incentives related to the more recent/current events are present, may underlie such efforts, in order to enhance the presumed effects of an injury or disorder. As in the

detection of exaggeration of symptoms, there are measures for detection of *minimization*, or under-reporting, of symptoms within certain instruments (Brown & Sellbom, 2020). When minimization of past symptoms is present, examiners should carefully evaluate self-report to determine whether malingering or other factors, such as stigma, potentially linked to cultural context, are contributing to the minimization of symptoms (Crichton et al., 2017).

Examiners should choose SVTs with documented reliability and validity and be aware of the sensitivity and specificity of their measures. SVTs with documented discriminatory ability and current interpretive guidelines relevant to the clinical condition of interest should be utilized. Ideally, SVTs should also provide non-redundant information and have validated cutoffs using known-groups (i.e. specificity  $\geq .90$ ; Sherman et al., 2020). Measures that provide quantification of validity relevant to the self-reported concerns/claimed disabilities are preferred. Scores on validity scales for feigned mental illness are not necessarily informative for self-reported cognitive or somatic complaints, consistent with the need to use multiple SVT measures. Effective practitioners remain up to date on the most recent advances on SVTs/self-reported validity analyses, as subsequent research with validity scales/tests often requires modification of interpretation.

On a final note, SVT data indicate whether the test taker is reporting symptoms in a credible manner. However, thoughtful practitioners know that simply reporting psychiatric symptoms credibly does not mean that an examinee has such symptoms at the time of testing. For example, individuals with a history of PTSD may well be able to report historically accurate PTSD symptoms, even if they no longer are experiencing such symptoms, and thereby appear to be presenting in a valid manner on test validity scales. The decision-making process requires that practitioners use additional history and case information to arrive at sound conclusions, including consideration of whether symptoms may have changed or improved over time.

### ***Integrated analysis of reported symptoms***

Having completed comprehensive record reviews (with data comparisons when available), conducted a thorough interview of the examinee (and collateral informants, when appropriate), and completed comprehensive assessment with appropriate SVT measures, examiners integrate their findings across these sources, including the neuropsychological test results and PVT data, for the purposes of determining the validity and interpretability of examinees' symptom reports. When symptom expression is judged to be consistent with the context of an extant disorder, which requires consideration of history and related other information, *and* when valid SVT data are present, examinee self-report may be considered to be valid and interpretable. If, however, assessment findings identify atypical or implausibly extreme severity in the context of (a) the medical record, (b) fact patterns of the case, (c) in-vivo examination and/or (d) SVT performance, the examinee's symptom report may be considered noncredible and uninterpretable. Guidance regarding SVT interpretation has been offered in several peer-reviewed publications and examiners are encouraged to be familiar with relevant, recent interpretative guidelines (e.g. Boone, 2017; Chafetz et al., 2020; Hall & Ben-

Porath, 2021). Similarly, when there is evidence of ‘compelling inconsistencies’ in the examinee’s presentation, such as differences in demeanor and presentation when being observed and when not, or inconsistencies between the medical record and observations, or inconsistencies between SVTs and observations, among others, then such inconsistencies should be considered to be part of an overall *invalid presentation* (Bianchini et al., 2005; Sherman et al., 2020).

On occasion, an examinee may be considered to have a credible mental health or medical disorder, but SVT results nevertheless indicate significant symptom overreporting/exaggeration that is inconsistent with the severity of the disorder. In such cases, the examiner should report such inconsistencies.

### *Recommendations for practitioners*

- Self-report inventories that do not contain a means for determining validity regarding symptom exaggeration/feigning and minimization/denial should not be used in isolation to make diagnostic or prognostic decisions. Self-report measures in the absence of objective SVT data are purely subjective and therefore are of unknown reliability and validity. The need for objective and valid data is crucial in both forensic and clinical assessment contexts.
- SVTs should be used in order to optimize confidence in self-report data collected from examinees.
- Failure on PVTs does not necessarily predict failure of SVTs; the converse is also true.
- SVTs should be domain specific (i.e. able to address specific claimed symptoms, whether somatic, psychological, or cognitive), with the understanding that symptom reporting may be highly correlated across domains and may be mutually informative.
- Failure on SVTs occurs when over-report or atypical report is present, which can happen within the context of evaluating individuals with any cognitive, somatic, or psychological disorder. Notably, the concern is not that a genuine disorder of these types causes invalidity, but that invalid self-report by an examinee may falsely suggest that the purported disorder is present, when it is not.
- Prior records may have been based on inaccurate/invalid self-report if SVTs were not administered and/or carefully considered as part of the prior evaluation.
- SVT results of examinees should be evaluated in terms of *discrepancy from normal expectations*. The presence of mild, moderate, and/or severe over-reporting of symptoms in an examinee’s assessment dataset should inform the ultimate conclusions/opinions of the examiner. Evidence of invalidity and/or response bias is not *necessarily* synonymous with malingering (i.e. over-reporting of symptoms may be relatively mild and not associated with an external incentive, leading to an impression of *mildly* exaggerated symptom reports, rather than frank malingering).
- When an examinee’s responses on SVTs result in extreme scale elevations, particularly on multiple scales/tests, and when the history/medical record is incompatible with severe psychopathology, especially when external incentives are known to be present or possibly present, examiners are encouraged to consider a determination

of malingering, utilizing contemporary criteria, when appropriate (e.g. Sherman et al., 2020). Consideration of the totality of the examinee's presentation is always essential, as a determination of malingering remains a complex, multidimensional process dependent on the entire record (Lippa, 2018). Additionally, exaggeration and/or malingering may coexist with genuine psychological, somatic, and cognitive disorders; they are not mutually exclusive (Merten & Merckelbach, 2013; Boone, 2017).

### *Recommendations for future research*

- There remains a need for development and validation of new SVTs, including free-standing measures and embedded measures within current self-report symptom measures. Validity assessment of self-reported functional impairment, which can be feigned, would also be valuable (Bryant et al., 2018; Suhr et al., 2017). Additional development and validation of SVTs for use with pediatric populations is clearly needed and strongly encouraged (Kirk et al., 2020).
- Ongoing validation and development of new measures that identify invalid collateral report is needed.
- General research in psychopathology, and in the development of new SVT scales and instruments, should be undertaken with research participants who themselves have passed SVTs that are appropriate for the individual's background. Diagnostic assessment of such research participants, whether recruited as patients or proxies, must include validation of their self-reported symptoms. Only by so doing can researchers, and ultimately examiners, have confidence in their instruments and their data.

### *Research design and statistical issues*

#### *Key points reaffirmed from the 2009 consensus conference statement*

In the original consensus conference statement (Heilbronner et al., 2009), three key areas were identified and discussed:

- **Research design.** The 2009 consensus conference statement endorsed the use of performance and symptom validity investigations that included analog simulation and criterion or "known groups" designs. Analog simulation designs provide a practical and cost-effective means for examining "proof of concept" for new tests. These allow for a comparison between a non-injured group, which is known to be feigning on measures of performance or symptom report, with that of an injured group with bona fide neurologic disorders such as a severe traumatic brain injury. The primary limitation of this design is one of generalizability. The strength of a known-groups design is the clinical relevance provided by including persons with real-world incentives to malingering or exaggerate their complaints or functional disability. The major methodological challenge is developing appropriate external criteria for defining response bias.

- **Single versus multiple indicators.** The 2009 consensus conference statement recommended more research investigating methods that combined multiple stand-alone tests and/or embedded measures. Relevant issues included: unit weighting approaches, logistic regression, multivariable composites, model over-fitting, and Bayesian model averaging. Whether using a multivariable composite or a single test, neuropsychologists were advised against relying on a single test with a fixed cut score. Because PVT and SVT diagnostic decisions occur in different contexts, the relative costs of false positive and false negative errors are not constant across situations. Raising or lowering a test's cut score will increase or decrease the test's sensitivity and specificity in an inverse fashion; when sensitivity is increased, specificity decreases. Consequently, investigators, journal editors, and test publishers were strongly encouraged to provide a broad range of cut scores with their respective diagnostic test statistics, including sensitivity, specificity, and likelihood ratios. Validation and cross-validation were also recommended for all measures and score combinations.
- **Diagnostic statistics.** This section discussed sensitivity, specificity, and how these statistics, in combination with the prevalence (i.e. base rate) of invalid performance, defined positive and negative predictive power. Also discussed were the benefits of using likelihood ratios (LR) and Receiver Operating Characteristic (ROC)/Area Under the Curve (AUC). AUC is particularly useful for comparing the diagnostic efficiency of individual PVTs to one another because it allows consideration of the full range of scores at every possible cut point. There are several interpretations of AUC: (1) the average value of sensitivity for all possible values of specificity; (2) the average value of specificity for all possible values of sensitivity; and, (3) the probability that a randomly selected person with the condition of interest has a test result indicating greater suspicion of that condition than that of a randomly chosen person without the condition of interest (Zhou et al., 2002). An AUC of .50 provides no diagnostic information, whereas an AUC of .70 to .80 shows acceptable discrimination, an AUC of .80 to .90 shows excellent discrimination, and  $AUC \geq .90$  shows outstanding discrimination (Hosmer & Lemeshow, 2000). AUC is related to both Cohen's *d* and Pearson *r* (Rice & Harris, 2005).

### *New consensus considerations on research design and statistical issues*

- **Research design.** Neuropsychological validity assessment research methods in the 2020s should include consideration of generalizability, or limitations, regarding such factors as ethnic/cultural diversity, language, and education. By comparison, there are fewer studies of normed and validated PVTs and SVTs for ethnic minorities, which creates an unknown potential for disparities when providing competent neuropsychological services to diverse populations. To be sure, research in this area can be challenging due to the combination of high heterogeneity in typically relatively small research samples. For example, in the U.S., there are about 20.4 million Asian-Americans, comprised of 19 specific ethnicities, that account for about 5.4% of the population (Lopez et al., 2017). Culture, sociohistorical experience, and

languages can differ for each ethnic group, and there can be significant within group differences in age, generation, acculturation, immigration history, level and quality of education, and English proficiency. Such factors can impact performance on neuropsychological tests (cf. review by Fujii, 2018) by impacting test validity, as well as an examinee's comfort with the testing situation, ability to respond to test items, and opportunities for learning (American Education Research Association et al., 2014). Despite the complexity of conducting such research, in alignment with the AACN Relevance 2050 initiative (Relevance 2050 Initiative - AACN (theaacn.org)), PVT and SVT studies with ethnic minorities are likely to be productive and important (cf. Robles et al., 2015).

When validating measures designed to detect performance and symptom validity, defining the valid and invalid criterion groups is a critical issue. For development of embedded/derived measures of performance validity, it is essential that PVTs identify performances that are atypical in either pattern or level of performance compared to the features characteristic of bona fide clinical impairment. For SVTs, one must identify patterns of symptom complaint that are atypical compared to those produced by both normal functioning individuals with similar cultural backgrounds and individuals with legitimate clinical disorders. For freestanding PVTs, it is critical that the test appears to be a measure of an actual neuropsychological ability, such as memory or attention, but is sufficiently easy to be performed normally by persons with genuine neurologic disorders. The key issue is not to confound true inability to perform the task with feigned inability.

Given the points made previously, great care is required for group formation. Validity assessment research requires a clinical comparison group that is characterized by a condition known to produce less effective neuropsychological performance and have a strong likelihood under most circumstances to manifest complaints of cognitive and/or psychological difficulties. Preferably, PVT and SVT research data on a clinical comparison group are obtained from individuals examined in the absence of external incentives. If external incentives are present, participants should be tested for evidence of valid responding. This creates a potential problem in that researchers might exclude individuals who fail a PVT because they lack the ability to earn a passing score. Such an error can inflate specificity, leading to an increase in false positive errors in subsequent applications of the PVT. These issues must be considered when researchers create samples for study using a known-groups design. Additionally, it is critical that there is a genuine clinical disorder comparison group in simulation research. With the exception of early validity investigations during development of a novel test application, comparisons made only between normal, uninjured individuals instructed to feign impairment and normal, uninjured individuals instructed to give valid performance/symptom reports are of questionable utility.

Data obtained from a study examining simulation versus genuine clinical disorder should be cross validated with an invalid performance criterion group. Specification of criteria for the invalid performance group should include multiple indicators of invalid performance. If this occurs in the absence of external incentive, this becomes a performance validity investigation designed to determine the

operating characteristics of a new PVT, or to cross-validate the operating characteristics of a previously developed PVT. If this occurs in the context of external incentive, such as litigation and/or disability compensation, or avoidance/mitigation of criminal prosecution, this becomes an investigation of probable malingering if invalidity is based on multiple PVT failures or definite malingering if the invalid performance group is defined by significantly-worse than chance performance on a two-alternative forced choice PVT (Binder et al., 2014).

The nature of the clinical disorders comprising both the noncredible and credible groups in a criterion groups design is important. Relevant to the practice of clinical neuropsychology, the most commonly litigated injury is uncomplicated mild traumatic brain injury (mTBI; Sweet et al., 2013; Sweet et al., 2021a), which is not expected to produce lasting neuropsychological impairment (Carroll et al., 2014; Hung et al., 2014; Rohling et al., 2011). Thus, persons with uncomplicated mTBI who fail multiple PVTs in the context of external incentives are performing in an invalid manner. This group can then be compared to individuals with uncomplicated mTBI who have passed all PVTs (cf. Jones, 2016; Jones, 2013). Although an appropriate design, such an approach limits use of the PVT under investigation to comparable cases, which, in this instance, is individuals with uncomplicated mTBI (i.e. failing PVTs vs. passing PVTs). Additional research would be required to support PVT application to a different clinical group, such as individuals with complicated mTBI.

In a variation of this design, an mTBI group in litigation that has failed multiple PVTs can be contrasted with a TBI group with either moderate or severe TBI, not in litigation. The advantage is that the mTBI group failing PVTs is not expected to have continuing neuropsychological impairment, contrasted with those with moderate or severe TBI who are at increased risk of persistent clinical impairment. This allows better determination of the clinically atypical performances characterizing noncredible presentations. In other words, uncomplicated mTBI is not expected to produce continuing cognitive impairment, and when such individuals fail PVTs, they have performed in a manner atypical even for persons having experienced more severe TBI. An additional analysis that this design allows is the demonstration of a dose-response effect on standard neuropsychological tests, such as measures of processing speed and memory (e.g. a case involving severe TBI would show worse performance than a case involving moderate TBI; Rohling et al., 2003; Sweet et al., 2013). Since the embedded PVT under development should represent an atypical pattern (i.e. not seen in moderate or severe TBI), the dose-response effect for the PVT in the credible performance group should be either absent or smaller than that seen on standard neuropsychological tests.

A third type of design uses mixed clinical samples, which may include a variety of clinical disorders, including neurologic (e.g. TBI & stroke), psychiatric (e.g. major depressive disorder), or developmental (e.g. learning disability) disorders. Cases passing PVTs are characterized as valid, while those failing PVTs are classified as invalid. A key issue with this type of design is ensuring that the valid and invalid groups do not differ in the type of clinical disorder or severity of clinical impairment (or if so, that the invalid group is over-represented by less severe conditions,

such as mTBI), so that, if present, group differences are only attributable to invalid performance. This concern is precluded when comparison groups consist of individuals with a single diagnosed condition that is of comparable severity across individuals (e.g. major depressive disorder) or when, with a single disorder, severity can be identified quantitatively and its effect examined statistically (e.g. moderate and severe TBI).

A significant concern in criterion group formation is what should be done with cases that fail only one PVT, but not at worse-than-chance levels. Some studies have suggested that a single PVT failure occurs in approximately 40% of credible neuropsychology clinic patients without motive to feign (Victor et al., 2009), whereas a single PVT failure likely represents performance invalidity in 57% of patients with motive to feign (Schroeder et al., 2019). This suggests that determination as to whether patients failing a single PVT should be included in a credible comparison group depends on presence of external motive. Preliminary research has shown that when patients without motive to feign and who fail a single PVT out of several are retained in a credible sample, patients failing 1 versus 0 PVTs may not differ in mean performance on the PVT under investigation, and cut-off selection is not impacted by inclusion of patients failing one PVT (McCaul et al., 2018), arguing for the inclusion of these individuals.

As found in surveys by Martin et al. (2015) and Schroeder et al. (2016), the most commonly preferred approaches to assessment of invalid performance require either failure of 2 or more PVTs, or failure of a single PVT plus evidence of one indicator of non-credibility outside of validity testing. By contrast, Victor et al. (2009) reported a high frequency of single PVT failures in their credible performance clinical sample, with no motive to feign. This finding, however, differed from the results of Proto, Pastorek, Miller, Romesser, Sim, & Linck (2014), who reported reduced neuropsychological performance for TBI cases failing only one PVT, compared to those subjects who passed all PVTs, although this was a military veteran sample that may have had actual or perceived motive to feign. Some investigators have chosen to simply exclude cases with one PVT failure as “indeterminate” (Schroeder et al., 2019); others have ranked the likelihood of invalid test scores along a continuum (Bianchini et al., 2014; Jones, Ingram, & Ben-Porath, 2012). The advantages and disadvantages of exclusion and inclusion have been discussed and initially tested empirically (Schroeder et al., 2019), with additional research needed on this important topic.

An alternate approach to research related to use of multiple PVTs is to exclude the single-fail PVT group, develop the cutting scores or logistic regression for the new PVT under investigation based on discriminating fail  $\geq 2$  versus fail 0 PVTs, and then apply the diagnostic discrimination procedure to the single-fail group in a post-hoc manner to see where these persons would be classified. This solution provides a means of addressing spectrum bias (i.e. variable performance because of different patient mix), which may in this instance be caused by dropping single-fail cases from the credible and noncredible groups. Such an approach can potentially provide better delineation of the boundaries of performance validity classification accuracy (see Loring et al., 2016 for discussion). However, to date, there is a lack of

consensus as to the best approach to classify the single-fail cases, which in the view of the consensus panel requires additional research.

In summary, there is consensus that simulation and criterion groups (“known groups”) research designs continue to be applied effectively in research on validity assessment. Such investigations depend on a careful characterization of comparison groups to ensure that the invalid performance group is truly atypical and that the clinical comparison group has an increased likelihood of legitimate neuropsychological deficits. This approach increases confidence that cutoffs for the PVT under investigation are based on clinically atypical as opposed to legitimately impaired performance. Various strategies for defining valid and invalid assessment groups are discussed in the preceding section on Research Designs. For example, there is consensual support for the research strategy of contrasting the performance of individuals with uncomplicated mTBI who are litigating/compensation seeking and who fail multiple PVTs with the performance of a sample of persons with moderate or severe TBI, not in litigation.

- **Diagnostic statistics.** By design, PVTs prioritize specificity over sensitivity, which is reflected in meta-analytic findings. Vickery et al. (2001) reported a mean sensitivity of .56 and specificity of .95, and, subsequently, Sollman and Berry (2011) reported a mean sensitivity of .69 and specificity, of .90. These studies also demonstrate the well-known inverse relationship between sensitivity and specificity: as sensitivity increases, specificity decreases. The formula for Positive Predictive Value (PPV; the probability of having the condition of interest given a positive diagnostic test result) also illustrates the emphasis on specificity with TP as true positives and FP as false positives (i.e.  $1 - \text{specificity}$ ):

$$\text{PPV} = \left[ \frac{\text{TP}}{\text{TP} + \text{FP}} \right] \quad (1)$$

Improvements in diagnostic accuracy can be made by reducing false positive error, in order to maximize the accurate detection of invalid performance

The importance of minimizing the false positive rate is further demonstrated by converting the mean sensitivity and specificity values from the two meta-analyses to positive likelihood ratios, (LR+) defined as:

$$\text{LR+} = \left[ \frac{\text{Sensitivity}}{1 - \text{Specificity}} \right] \quad (2)$$

For the Vickery et al. (2001) data, this value is 11.2 (i.e.  $.56/.05$ ), indicating that a positive score on the PVTs in this investigation was 11 times more likely to come from the invalid performance than the valid performance group. For the Sollman and Berry (2011) data, this value is 6.9 (i.e.  $.69/.10$ ), or 7 times more likely to come from the invalid as opposed to the valid performance group.

Moreover, these LR+ values exceed the LR+ for discrimination of patients with neurologic impairment from normal control subjects, 5.28, based on the average impairment rating of the Halstead-Reitan Neuropsychological Battery (Heaton, Miller, Taylor, & Grant, 2004; Rohling et al., 2021).

Finally, note that LR+ multiplied by the base rate odds of the presence of non-credible performance yields the post-test odds of the presence of noncredible

performance. So, using a base rate of .40 of noncredible performance, the base rate odds become:

$$\text{Pre - test odds} = [.40/(1 - .40)] = 0.67 \quad (3)$$

Thus, when multiplied by the respective LR+ values, the result equals post-test odds of 7.5 for Vickery et al. (2001), and 4.6 for Sollman and Berry (2011). These post-test odds can then be converted to post-test probabilities of noncredible performance by the formula:

$$\text{Post - test probability} = [\text{odds}/(\text{odds} + 1)] \quad (4)$$

This yields probabilities of invalid performance of 0.88 for Vickery et al. (2001) and 0.82 for Sollman and Berry (2011).

Consensus continues that diagnostic statistics are important in the research and application of PVTs and SVTs. More specifically, these investigations demonstrate the importance of maintaining a low false positive rate to improve diagnostic accuracy. Current consensus also supports a false positive rate of .10 (specificity of .90) per PVT or SVT. This is supported by the Sollman and Berry (2011) meta-analysis yielding an average false positive of .10 per PVT, and by Larrabee et al. (2019) who varied the per-test false positive rate from 0.00 to 0.15, and found .10 to be the optimal rate. The Sherman et al. (2020) revision of the diagnostic criteria for malingered neuropsychological dysfunction also recommends a per-PVT and per SVT false positive rate of .10, as well as an aggregate false positive rate of .10 for use of combinations of multiple PVTs and SVTs.

- **Use of single versus multiple indicators.** The 2009 consensus conference statement recommended that practitioners use multiple indicators of performance and symptom validity, and in the intervening years this has, in fact, happened (Martin et al., 2015; Sweet et al., 2021b). Such recommendations, in conjunction with the explosion in PVT and SVT research over the past 10+ years, raised subsequent concerns regarding the best approach for combining multiple PVTs and SVTs in an individual assessment. A related concern is the potential for elevated false positive rates associated with use of multiple PVTs and SVTs.

As discussed in the original consensus conference statement, two basic approaches have been employed: A simple tally method with unit weighting, and more elaborate statistical models such as discriminant function analysis and logistic regression. The advantage of statistical models such as logistic regression is that PVT and SVT correlations are considered, as well as the potential to provide different weightings for more salient PVT and SVT measures. The disadvantage is the lack of a consensus core neuropsychological assessment battery, employing a common set of PVTs and SVTs that would allow derivation of a logistic regression equation.

The tally approach is dependent on the generalizability of study-specific results on performance contingencies of various PVT and SVT data (Larrabee, 2003; Victor et al., 2009). The lack of a core battery with common measures has led to various attempts to provide a general model for a combination of measures. These have

included linking/chaining of LR+ (Larrabee, 2008), LR+ and LR- (Bender & Frederick, 2018), and Monte Carlo simulation (Berthelson et al., 2013). These methods have been criticized for over-estimation of posterior probabilities when chaining of all possible LR+ and LR- combinations are considered (Larrabee et al., 2019), and for over-estimation of false positive errors by Monte Carlo simulation compared to values generated from actual clinical data (Davis & Millis, 2014; Larrabee, 2014). Overestimation of error rate using Monte Carlo analysis is likely due to skewed data in credible performance groups (Davis & Millis, 2014; Larrabee et al., 2019). This is because PVTs are criterion-referenced, not norm-referenced (Davis, 2018).

The Davis and Millis (2014) and Larrabee (2014) data support using a cutoff of  $\geq 2$  PVT or SVT failures when up to 7 to 9 measures are administered. The Sherman et al. (2020) MND criteria paper also advises a cut-off of  $\geq 2$  PVTs, (e.g. 2 of 7 failed), but recommends a ratio of PVTs failed to those administered when the number of PVTs administered is greater (e.g. 4 of 14 failed), to maintain a low false positive rate. While based on a coherent rationale, empirical confirmation is currently insufficient and considered necessary to ensure that specificity in fact remains at adequate levels. Additionally, PVT failure rates in non-feigning samples are highly likely to be non-linear, due to skewed score distributions (Larrabee et al., 2019), raising concerns regarding computing the ratio of number failed to number administered.

In summary, consensus continues regarding the need to utilize multiple PVTs and SVTs in validity assessment research. There is no specification as to the exact number of PVTs and SVTs that should be administered in individual cases, but there is consensus that multiple PVTs and SVTs should be employed. There is also consensus that increasing numbers of PVT failures most likely represent performance invalidity, but additional investigations are needed to precisely measure false positive rates when a high number of PVTs are failed. Ultimately, the determination of malingering is not solely based on tallies or ratios of PVT/SVT failure, but whether the information pertaining to the case meets criteria in the accepted multidimensional guidelines.

### *Future directions*

Concern regarding false positive identifications continues to be an important area of focus of PVT and SVT research. Investigators are urged to report the clinical characteristics of false positive cases; that is, invalid scores produced by credible cases performing similarly to noncredible cases on the PVT or SVT under investigation. With this, the clinician employing the PVT or SVT can determine if a particular case has a clinical history or other characteristics similar to those of false positive cases; if not, the case failing the validity measure may well be a true positive for invalid performance. False positive identifications on validity measures are associated with living in a residential facility with 24-hour supervision (Meyers & Volbrecht, 2003), severe TBI associated with prolonged coma (Larrabee, 2003), dementia with increasing false positive rate associated with increasing degree of dementia (Dean et al., 2009; Loring et al., 2016),

schizophrenia with comorbid significant cognitive impairment (Ruiz et al., 2020), and sometimes amnesic mild cognitive impairment (Loring et al., 2016; Schroeder & Martin, 2021c).

Procedures aimed at reducing the false positive rate in credible clinical samples have been suggested. Dean et al. (2009) offered adjustments for cutting scores for individual PVTs to maintain specificity at .90. Loring et al. (2016) found that false positive rates for early AD were 13% for  $RDS \leq 6$ , and 70% for Rey Auditory Verbal Learning Test (AVLT; Rey, 1964) Recognition  $\leq 9$ , but that combining these cutting scores ( $RDS \leq 6$  and AVLT Recognition  $\leq 9$ ) caused the false positive rate to fall to just 5% for AD and 1% for amnesic mild cognitive impairment.

Loring et al. (2016) also demonstrated that evidence of normal range performances on Trails B and on AVLT delayed free recall were associated with a decrease in false positive rates for RDS and AVLT Recognition. In other words, evidence of normal range performance on sensitive measures of memory and processing speed suggested sufficient native ability to perform validly on RDS and AVLT Recognition.

This information supports the need for additional investigations of PVT and SVT failure rate in clinical samples without external incentive (e.g. individuals not in litigation or seeking or already receiving disability compensation). This would not only lead to documentation of the neurologic, psychiatric, and developmental disorders known to be associated with PVT and SVT failure, but could also potentially lead to a means of reducing the false positive rate by combining patient-related variables (educational attainment, language proficiency, etc.) and injury/illness-related variables such as level of ability required to perform the PVTs under investigation.

Also recommended is continued investigation of the diagnostic accuracy of varying combinations of multiple PVTs and SVTs. Cross validation is important, both for clinical criterion groups (i.e. neurologic, psychiatric, and developmental) and for determination of the increase in false positive rates associated with multiple measures. In other words, at what point does the requirement of  $\geq 2$  PVT failures lead to an unacceptable false positive rate due to the use of 10, 15, or 20 PVTs and SVTs? Finally, PVT research also needs to focus on multiple domains of performance and symptom validity, which have been discussed in the PVT and SVT sections of the current consensus paper. Memory-based validity has been the most investigated domain; additional research is needed for ongoing PVT development in language, perception, sensorimotor, attention, processing speed, and problem-solving validity. SVT failure can occur with claims of psychosis, PTSD, depression, anxiety, and cognitive and somatic symptoms, which were discussed previously in the relevant sections in the present consensus paper.

### Summary of current consensus

That valid information is required in order to form accurate opinions regarding examinee presentation has been increasingly recognized and emphasized by clinical neuropsychologists. There is now a nearly universal acceptance among clinical neuropsychology postdoctoral trainees and practicing clinical neuropsychologists (Sweet et al., 2021a, 2021b) that proactive evaluation of validity, including reliance on

objective validity indicators, is an important aspect of assessment, regardless of whether the evaluation context is clinical or forensic. Beyond practitioner-level support for the opinion that all neuropsychological evaluations should proactively include validity measures, there continues to be substantial emphasis and support in the expert literature providing guidance regarding application of validity assessment procedures for clinicians in general, not just forensic practitioners (cf. Schroeder & Martin, 2021d).

Since the original 2009 consensus conference statement, relevant terminology has shifted away from broad and sometimes ill-fitting concepts, such as *effort*, and sharpened toward more fundamental concepts of *validity*, measured deliberately and with greater objective precision. Additionally, validity testing has been recategorized into performance-based PVTs and symptom-reported SVTs (Larrabee, 2012). The need for tighter research methods underlying related clinical procedures has also been emphasized (cf. Schroeder et al., 2019; Sherman et al., 2020), with evidence-based topics, such as use of multiple indicators, continuing to better inform practice, even as consensus is awaited on such topics. There continues to be substantial annual knowledge growth from relevant empirically-based research studies, with a continued burgeoning of relevant peer-reviewed research unmistakably evident in mainstream neuropsychology practice journals, evincing more published studies now addressing validity assessment in clinical than in forensic contexts (Suchy, 2019).

As was true in 2009, the sizeable subject matter of validity assessment is being energetically pursued by clinical neuropsychology practitioners and researchers and it will continue to evolve, with the net effect being that the guidance from the present consensus panel will inevitably require future updating. Keeping abreast of distinctions of concept and procedures with regard to identifying invalid presentations, and the subset that can even more accurately be depicted as malingering, requires an ongoing commitment from practitioners to maintain competencies on these important practice topics.

## Disclosure statement

All participants were asked to complete a current AACN conflict of interest statement. No conflicts were identified. Within this document, select specific tests are named, merely as examples.

## References

- Airaksinen, E., Larsson, M., & Forsell, Y. (2005). Neuropsychological functions in anxiety disorders in population-based samples: Evidence of episodic memory dysfunction. *Journal of Psychiatric Research, 39*(2), 207–214. <https://doi.org/10.1016/j.jpsychires.2004.06.001>
- American Academy of Clinical Neuropsychology (AACN) (2007). American Academy of Clinical Neuropsychology (AACN) Practice Guidelines for neuropsychological assessment and consultation. *The Clinical Neuropsychologist, 21*, 209–231.
- American Education Research Association, American Psychological Association, & the National Council on Measurement in Education (2014). *Standards for educational and psychological testing* (2nd ed.). American Education Research Association.
- American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text revision). American Psychiatric Association.
- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). American Psychiatric Association.

- American Psychological Association. (2013). Specialty guidelines for forensic psychology. *The American Psychologist*, 68, 7–19.
- Andrikopoulos, J. (2018). Clinical assessment of posttraumatic stress disorder. In J. E. Morgan & J. H. Ricker (Eds.), *Textbook of clinical neuropsychology* (2nd ed., pp. 757–791). Routledge/Taylor & Francis.
- Andrikopoulos, J., & Greiffenstein, M. F. (2012). Something to talk about? The status of post-traumatic stress disorder in clinical neuropsychology. In G. J. Larrabee (Ed.), *Forensic neuropsychology: A scientific approach* (pp. 365–400). Oxford University.
- Ardolf, B. R., Denney, R. L., & Houston, C. M. (2007). Base rates of negative response bias and Malingered Neurocognitive Dysfunction among criminal defendants referred for neuropsychological evaluation. *The Clinical Neuropsychologist*, 21(6), 899–916. <https://doi.org/10.1080/13825580600966391>
- Bass, C., & Wade, D. T. (2019). Malingering and factitious disorder. *Practical Neurology*, 19(2), 96–105. <https://doi.org/10.1136/practneurol-2018-001950>
- Belanger, H. G., Tate, D., & Vanderploeg, R. D. (2018). Concussion and mild traumatic brain injury. In J. E. Morgan & Ricker, J. H. (Eds.), *Textbook of clinical neuropsychology* (2nd ed., pp. 411–448). Routledge/Taylor & Francis.
- Bender, S. D., & Frederick, F. (2018). Neuropsychological models of feigned cognitive deficits. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (4th ed., pp. 42–60). Guilford.
- Bender, S. D., & Matuszewicz, M. (2013). PCS, iatrogenic symptoms, and malingering following concussion. *Psychological Injury and Law*, 6(2), 113–121. <https://doi.org/10.1007/s12207-013-9156-9>
- Ben-Porath, Y. S., Graham, J. R., & Tellegen, A. (2009). Development, research findings, and interpretive recommendations (2009). *MMPI-2 symptom validity (FBS) scale*. University of Minnesota.
- Ben-Porath, Y. S., & Tellegen, A. (2008). *MMPI-2-RF: Minnesota multiphasic personality inventory-2 restructured form*. Pearson.
- Berthelson, L., Mulchan, S. S., Odland, A. P., Miller, L. J., & Mittenberg, W. (2013). False positive diagnosis of malingering due to the use of multiple effort tests. *Brain Injury*, 27(7-8), 909–916. <https://doi.org/10.3109/02699052.2013.793400>
- Bianchini, K. J., Aguerrevere, L. E., Curtis, K. L., Roebuck-Spencer, T. M., Frey, F. C., Greve, K. W., & Calamia, M. (2018). Classification accuracy of the Minnesota Multiphasic Personality Inventory-2 (MMPI-2)-Restructured form validity scales in detecting malingered pain-related disability. *Psychological Assessment*, 30(7), 857–869. <https://doi.org/10.1037/pas0000532>
- Bianchini, K. J., Aguerrevere, L. E., Guise, B. J., Ord, J. S., Etherton, J. L., Meyers, J. E., Soignier, R. D., Greve, K. W., Curtis, K. L., & Bui, J. (2014). Accuracy of the Modified Somatic Perception Questionnaire and Pain Disability Index in the detection of malingered pain-related disability in chronic pain. *The Clinical Neuropsychologist*, 28(8), 1376–1394. <https://doi.org/10.1080/13854046.2014.986199>
- Bianchini, K. J., Greve, K. W., & Glynn, G. (2005). On the diagnosis of malingered pain-related disability: Lessons from cognitive malingering research. *The Spine Journal*, 5(4), 404–417. <https://doi.org/10.1016/j.spinee.2004.11.016>
- Binder, L. M., Larrabee, G. J., & Millis, S. R. (2014). Intent to fail: Significance testing of forced choice test results. *The Clinical Neuropsychologist*, 28(8), 1366–1375. <https://doi.org/10.1080/13854046.2014.978383>
- Boone, K. B. (2007). A reconsideration of the Slick et al. (1999) criteria for malingered neurocognitive dysfunction. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective*. Guilford.
- Boone, K. B. (2009a). Fixed belief in cognitive dysfunction despite normal neuropsychological scores: Neurocognitive Hypochondriasis? *The Clinical Neuropsychologist*, 23(6), 1016–1036. <https://doi.org/10.1080/13854040802441135>

- Boone, K. B. (2009b). The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *The Clinical Neuropsychologist*, 23(4), 729–741. <https://doi.org/10.1080/13854040802427803>
- Boone, K. B. (2011). Somatoform disorders, factitious disorder, and malingering. In M. Schoenberg, *The little black book of neuropsychology: A Syndrome-based approach*. Springer.
- Boone, K. B. (2013). *Clinical practice of forensic neuropsychology: An evidence-based approach*. Guilford.
- Boone, K. B. (2017). Self-deception in somatoform conditions: Differentiating between conscious and nonconscious symptom feigning. In K. B. Boone (Ed.), *Neuropsychological evaluation of somatoform and other functional somatic conditions: Assessment primer* (pp. 3–42). Routledge/Taylor & Francis.
- Boone, K. B. (2018). Assessment of neurocognitive performance validity. In J.E. Morgan & J.H. Ricker (Eds.), *Textbook of clinical neuropsychology* (2nd ed., pp. 39–50). Routledge/Taylor & Francis.
- Brooks, B. L., Ploetz, D. M., & Kirkwood, M. W. (2016). A survey of neuropsychologists' use of validity tests with children and adolescents. *Child Neuropsychology*, 22(8), 1001–1020.
- Broshek, D. K., Marco, A. P., & Freeman, J. R. (2015). A review of post-concussion syndrome and psychological factors associated with concussion. *Brain Injury*, 29(2), 228–237. <https://doi.org/10.3109/02699052.2014.974674>
- Brown, R. (2004). Psychological mechanisms of medically unexplained symptoms: An integrative conceptual model. *Psychological Bulletin*, 130(5), 793–812. <https://doi.org/10.1037/0033-2909.130.5.793>
- Brown, T. A., & Sellbom, M. (2020). The utility of the MMPI-2-RF validity scales in detecting underreporting. *Journal of Personality Assessment*, 102(1), 66–74. <https://doi.org/10.1080/00223891.2018.1539003>
- Bryant, A. M., Lee, E., Howell, A., Morgan, B., Cook, C. M., Patel, K., Menatti, A., Clark, R., Buelow, M. T., & Suhr, J. A. (2018). The vulnerability of self-reported disability measures to malingering: A simulated ADHD study. *The Clinical Neuropsychologist*, 32(1), 109–118. <https://doi.org/10.1080/13854046.2017.1346145>
- Bush, S. S., Connell, M. A., & Denney, R. L. (2020). *Ethical practice in forensic psychology: A guide for mental health professionals* (2nd ed.). American Psychological Association.
- Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., Reynolds, C. R., & Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity NAN policy & planning committee. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 20(4), 419–426. <https://doi.org/10.1016/j.acn.2005.02.002>
- Carone, D. A. (2018). Medical and psychological iatrogenesis in neuropsychological assessment. In Morgan, J.E. & Ricker, J.H. (Eds.) *Textbook of clinical neuropsychology* (2nd ed., pp. 1108–1031). Routledge/Taylor & Francis.
- Carroll, L. J., Cassidy, J. D., Cancelliere, C., Côté, P., Hincapié, C. A., Kristman, V. L., Holm, L. W., Borg, J., Nygren-de Bousard, C., & Hartvigsen, J. (2014). Systematic review of the prognosis after mild traumatic brain injury in adults: cognitive, psychiatric, and mortality outcomes: Results of the international collaboration on mild traumatic brain injury prognosis. *Archives of Physical Medicine and Rehabilitation*, 95(3 Suppl), S152–S173. <https://doi.org/10.1016/j.apmr.2013.08.300>
- Carson, A., Hallett, M., & Stone, J. (2016). Assessment of patients with functional neurologic disorders. (pp. 169–188) In (Eds). M. Hallett, J. Stone, and A. Carson, *Functional Neurologic Disorders Vol. 130 of Handbook of Clinical Neurology* (3rd ed.). Elsevier. <https://doi.org/10.1016/B978-0-12-801772-2.00015-1>
- Chafetz, M. D. (2008). Malingering on the Social Security Disability Consultative Exam: Predictors and base rates. *The Clinical Neuropsychologist*, 22(3), 529–546. <https://doi.org/10.1080/13854040701346104>

- Chafetz, M. D. (2021). Deception is different: Negative validity test findings do not provide “evidence” for “good effort. *The Clinical Neuropsychologist*, 35. <https://doi.org/10.1080/13854046.2020.1840633>
- Chafetz, M. D., Bauer, R. M., & Haley, P. S. (2020). The other face of illness-deception: Diagnostic criteria for factitious disorder with proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, 34(3), 454–476. <https://doi.org/10.1080/13854046.2019.1663265>
- Chafetz, M., & Dufrene, M. (2014). Malingering by proxy: Need for child protective services and guidance for reporting. *Child Abuse & Neglect*, 38(11), 1755–1765. <https://doi.org/10.1016/j.chiabu.2014.08.015>
- Clark, H. C., Martin, P. K., Okut, H., & Schroeder, R. W. (2020). A systematic review and meta-analysis of the utility of the Test of Memory Malingering in pediatric examinees. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 35(8), 1312–1322. <https://doi.org/10.1093/arclin/aca0075>
- Colbert, A. M., Maxwell, E. C., & Kirkwood, M. W. (2021). Validity assessment in pediatric populations. In K. B. Boone (Ed.) *Assessment of feigned cognitive impairment: A neuropsychological perspective* (2nd ed.). Guilford.
- Cotrena, C., Branco, L. D., Ponsoni, A., Shansis, F. M., Kochhann, R., & Fonseca, R. P. (2017). The predictive role of daily cognitive stimulation on executive functions in bipolar disorder. *Psychiatry Research*, 252, 256–261. <https://doi.org/10.1016/j.psychres.2017.03.011>
- Crighton, A. H., Marek, R. J., Dragon, W. R., & Ben-Porath, Y. S. (2017). Utility of the MMPI-2-RF validity scales in detection of simulated underreporting: Implications of incorporating a manipulation check. *Assessment*, 24(7), 853–864. <https://doi.org/10.1177/1073191115627011>
- Critchfield, E., Soble, J. R., Marceaux, J. C., Bain, K. M., Chase Bailey, K., Webber, T. A., Alex Alverson, W., Messerly, J., Andrés González, D., & O'Rourke, J. J. F. (2019). Cognitive impairment does not cause invalid performance: Analyzing performance patterns among cognitively unimpaired, impaired, and noncredible participants across six performance validity tests. *The Clinical Neuropsychologist*, 33(6), 1083–1101. <https://doi.org/10.1080/13854046.2018.1508615>
- Dandachi-FitzGerald, B., Merckelbach, H., & Ponds, R. W. (2017). Neuropsychologists' ability to predict distorted symptom presentation. *Journal of Clinical and Experimental Neuropsychology*, 39(3), 257–264.
- Davis, J. J. (2018). Performance validity in older adults: Observed versus predicted false positive rates in relation to number of tests administered. *Journal of Clinical and Experimental Neuropsychology*, 40(10), 1013–1021. <https://doi.org/10.1080/13803395.2018.1472221>
- Davis, J. J., & Millis, S. R. (2014). Examination of performance validity test failure in relation to number of tests administered. *The Clinical Neuropsychologist*, 28(2), 199–214. <https://doi.org/10.1080/13854046.2014.884633>
- Dean, A. C., Victor, T. L., Boone, K. B., Philpott, L. M., & Hess, R. A. (2009). Dementia and effort test performance. *The Clinical Neuropsychologist*, 23(1), 133–152. <https://doi.org/10.1080/13854040701819050>
- Delis, D. C., & Wetter, S. R. (2007). Cogniform disorder and cogniform condition: Proposed diagnoses for excessive cognitive symptoms. *Archives of Clinical Neuropsychology*, 22(5), 589–604. <https://doi.org/10.1016/j.acn.2007.04.001>
- Demakis, G. J., Gervais, R. O., & Rohling, M. L. (2008). The effect of failure on cognitive and psychological symptom validity tests in litigants with symptoms of post-traumatic stress disorder. *The Clinical Neuropsychologist*, 22(5), 879–895. <https://doi.org/10.1080/13854040701564482>
- Denney, R. L. (2008). Negative response bias and malingering during neuropsychological assessment in criminal forensic settings. In R. L. Denney & J. P. Sullivan (Eds.), *Clinical neuropsychology in the criminal forensic setting* (pp. 91–134). Guilford.
- DeRight, J., & Carone, D. A. (2015). Assessment of effort in children: A systematic review. *Child Neuropsychology: a Journal on Normal and Abnormal Development in Childhood and Adolescence*, 21(1), 1–24. <https://doi.org/10.1080/09297049.2013.864383>
- Dikmen, S., Machamer, J., Fann, J. R., & Temkin, N. R. (2010). Rates of symptom reporting following traumatic brain injury. *Journal of the International Neuropsychological Society*, 16(3), 401–411. <https://doi.org/10.1017/S1355617710000196>

- Dorociak, K. E., Schulze, E. T., Piper, L. E., Molokie, R. E., & Janeczek, J. K. (2018). Performance validity testing in a clinical sample of adults with sickle cell disease. *The Clinical Neuropsychologist*, 32(1), 81–97. <https://doi.org/10.1080/13854046.2017.1339830>
- East-Richard, C., R. Mercier, A., Nadeau, D., & Cellard, C. (2020). Transdiagnostic neurocognitive deficits in psychiatry: A review of meta-analyses. *Canadian Psychology/Psychologie Canadienne*, 61(3), 190–214. <https://doi.org/10.1037/cap0000196>
- Edmonds, E. C., Delano-Wood, L., Galasko, D. R., Salmon, D. P., & Bondi, M. W. (2014). Subjective cognitive complaints contribute to misdiagnosis of mild cognitive impairment. *Journal of the International Neuropsychological Society: JINS*, 20(8), 836–847. <https://doi.org/10.1017/S135561771400068X>
- Faust, D., Hart, K., Guilmette, T. J., & Arkes, H. R. (1988). Neuropsychologists' capacity to detect adolescent malingerers. *Professional Psychology: Research and Practice*, 19(5), 508–515. <https://doi.org/10.1037/0735-7028.19.5.508>
- French, L. M., Cernich, A. N., & Howe, L. L. (2018). Military service-related traumatic brain injury. In J.E. Morgan & J.H. Ricker (Eds.), *Textbook of clinical neuropsychology* (2nd ed., pp. 792–822). Routledge/Taylor & Francis.
- Fujii, D. E. (2017). *Conducting a culturally informed neuropsychological evaluation*. American Psychological Association.
- Fujii, D. E. (2018). Developing a cultural context for conducting a neuropsychological evaluation with a culturally diverse client: the ECLECTIC framework. *The Clinical Neuropsychologist*, 32(8), 1356–1337. <https://doi.org/10.1080/13854046.2018.1435826>
- Gass, C. S., & Curiel, R. E. (2011). Test anxiety in relation to measures of cognitive and intellectual functioning. *Archives of Clinical Neuropsychology*, 26(5), 396–404. <https://doi.org/10.1093/arclin/acr034>
- Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Sellbom, M. (2010). Incremental validity of the MMPI-2-RF over-reporting scales and RBS in assessing the veracity of memory complaints. *Archives of Clinical Neuropsychology*, 25(4), 274–284.
- Gervais, R. O., Rohling, M. L., Green, P. W., & Ford, W. (2004). A comparison of WMT, CARB, and TOMM failure rates in non-head injury disability claimants. *Archives of Clinical Neuropsychology*, 19(4), 475–488. <https://doi.org/10.1016/j.acn.2003.05.001>
- Goldfried, M. R. (2013). Evidence-based treatment and cognitive-affective-relational-behavior-therapy. *Psychotherapy (Chicago, Ill.)*, 50(3), 376–380. <https://doi.org/10.1037/a0032158>
- Goodwin, B. E., Sellbom, M., & Arbisi, P. A. (2013). Posttraumatic stress disorder in veterans: The utility of the MMPI-2-RF validity scales in detecting overreported symptoms. *Psychological Assessment*, 25(3), 671–678. <https://doi.org/10.1037/a0032214>
- Gottfried, E., & Glassmire, D. (2016). The relationship between psychiatric and cognitive symptom feigning among forensic inpatients adjudicated incompetent to stand trial. *Assessment*, 23(6), 672–682. <https://doi.org/10.1177/1073191115599640>
- Green, P. (2008). Questioning common assumptions about depression. In J. E. Morgan & J. J. Sweet (Eds.), *Neuropsychology of malingering casebook* (pp. 132–144). Psychology Press/Taylor & Francis.
- Green, P., & Merten, T. (2013). Noncredible explanations of noncredible performance on symptom validity tests. In D. A. Carone & S. S. Bush (Eds.), *Mild traumatic brain injury: Symptom validity assessment and malingering* (pp. 73–100). Springer.
- Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment*, 6(3), 218–224.
- Greiffenstein, M. F., Fox, D., & Lees-Haley, P. R. (2007). The MMPI-2 fake bad scale in detection of noncredible brain injury claims. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 210–235). Guilford.
- Greve, K. W., Bianchini, K. J., & Brewer, S. T. (2013). The assessment of performance and self-report validity in persons claiming pain-related disability. *The Clinical Neuropsychologist*, 27(1), 108–137. <https://doi.org/10.1080/13854046.2012.739646>

- Greve, K. W., Bianchini, K. J., & Brewer, S. T. (2018). Pain and pain-related disability. In J.E. Morgan & J.H. Ricker (Eds.), *Textbook of clinical neuropsychology* (2nd ed., pp. 823–845). Routledge/Taylor & Francis Group.
- Greve, K. W., Ord, J. S., Bianchini, K. J., & Curtis, K. L. (2009). Prevalence of malingering in patients with chronic pain referred for psychologic evaluation in a medico-legal context. *Archives of Physical Medicine and Rehabilitation*, *90*(7), 1117–1126. <https://doi.org/10.1016/j.apmr.2009.01.018>
- Guilmette, T. J., Hart, K. J., Giuliano, A. J., & Leininger, B. E. (1994). Detecting simulated memory impairment: Comparison of the Rey Fifteen-Item Test and the Hiscock forced-choice procedure. *Clinical Neuropsychologist*, *8*(3), 283–294. <https://doi.org/10.1080/13854049408404135>
- Hall, J. T., & Ben-Porath, Y. S. (2021). The MMPI-2-RF validity scales: An overview of research and applications. In R. W. Schroeder & P. K. Martin (Eds.), *Validity assessment in clinical neuropsychological practice*. Guilford. (in press).
- Harp, J. P., Jasinski, L. J., Shandera-Ochsner, A. L., Mason, L. H., & Berry, D. T. (2011). Detection of malingered ADHD using the MMPI-2-RF. *Psychological Injury and Law*, *4*(1), 32–43. <https://doi.org/10.1007/s12207-011-9100-9>
- Harrison, A. G., & Armstrong, I. T. (2016). Development of a symptom validity index to assist in identifying ADHD symptom exaggeration or feigning. *The Clinical Neuropsychologist*, *30*(2), 265–283. <https://doi.org/10.1080/13854046.2016.1154188>
- Harrison, A. G., Flaro, L., & Armstrong, I. (2015). Rates of effort test failure in children with ADHD: An exploratory study. *Applied Neuropsychology. Child*, *4*(3), 197–210. <https://doi.org/10.1080/21622965.2013.850581>
- Harrison, A. G., Lee, G. J., & Suhr, J. A. (2021). Use of performance validity tests and symptom validity tests in assessment of specific learning disorders and attention-deficit/hyperactivity disorder. In K. B. Boone, (Ed.) *Assessment of feigned cognitive impairment: A neuropsychological perspective* (2nd ed.). Guilford.
- Heaton, R. K., Miller, S. W., Taylor, M. J., & Grant, I. (2004). Revised comprehensive norms for an expanded Halstead-Reitan Battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults, professional manual. *Psychological Assessment Resources*; Odessa, FL.
- Heaton, R. K., Smith, H. H., Lehman, R. A. W., & Vogt, A. T. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology*, *46*(5), 892–900. <https://doi.org/10.1037//0022-006x.46.5.892>
- Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants 1. (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, *23*(7), 1093–1129. <https://doi.org/10.1080/13854040903155063>
- Henry, G. K., Heilbronner, R. L., Suhr, J., Gornbein, J., Wagner, E., & Drane, D. L. (2018). Illness perceptions predict cognitive performance validity. *Journal of the International Neuropsychological Society: JINS*, *24*(7), 735–745. <https://doi.org/10.1017/S1355617718000218>
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). John Wiley & Sons. <https://doi.org/10.1002/0471722146>
- Hung, R., Carroll, L. J., Cancelliere, C., Côté, P., Rumney, P., Keightley, M., Donovan, J., Stålnacke, B.-M., & Cassidy, J. D. (2014). Systematic review of the clinical course, natural history, and prognosis for pediatric mild traumatic brain injury: results of the International Collaboration on Mild Traumatic Brain Injury Prognosis. *Archives of Physical Medicine and Rehabilitation*, *95*(3 Suppl), S174–S191. <https://doi.org/10.1016/j.apmr.2013.08.301>
- Hurtubise, J. L., Scavone, A., Sagar, S., & Erdodi, L. A. (2017). Psychometric markers of genuine and feigned neurodevelopmental disorders in the context of applying for academic accommodations. *Psychological Injury and Law*, *10*(2), 121–137. <https://doi.org/10.1007/s12207-017-9287-5>
- Ingram, P. B., & Ternes, M. S. (2016). The detection of content-based invalid responding: a meta-analysis of the MMPI-2-Restructured Form's (MMPI-2-RF) over-reporting validity scales. *The Clinical Neuropsychologist*, *30*(4), 473–496. <https://doi.org/10.1080/13854046.2016.1187769>

- Iverson, G. L. (2007). Identifying exaggeration and malingering. *Pain Practice: The Official Journal of World Institute of Pain*, 7(2), 94–102. <https://doi.org/10.1111/j.1533-2500.2007.00116.x>
- Iverson, G. L. (2012). A biopsychosocial conceptualization of poor outcome from mild traumatic brain injury. In R. A. Vasterling, R. A. Bryant, & T. M. Keane (Eds). *PTSD and mild traumatic brain injury* (pp. 37–60). Guilford.
- Jones, A. (2013). Test of Memory Malingering: Cutoff scores for psychometrically defined malingering groups in a military sample. *The Clinical Neuropsychologist*, 27(6), 1043–1059. <https://doi.org/10.1080/13854046.2013.804949>
- Jones, A. (2016). Repeatable Battery for the Assessment of Neuropsychological Status: Effort index cutoff scores for psychometrically defined malingering groups in a military sample. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 31(3), 273–283. <https://doi.org/10.1093/arclin/acw006>
- Jones, A., Ingram, M. V., & Ben-Porath, Y. S. (2012). Scores on the MMPI-2-RF scales as a function of increasing levels of failure on cognitive symptom validity tests in a military sample. *The Clinical Neuropsychologist*, 26(5), 790–815. <https://doi.org/10.1080/13854046.2012.693202>
- Kalfon, T. B. O., Gal, G., Shorer, R., & Ablin, J. N. (2016). Cognitive functioning in fibromyalgia: the central role of effort. *Journal of Psychosomatic Research*, 87, 30–36. <https://doi.org/10.1016/j.jpsychores.2016.06.004>
- Kansner, R. J., Rapport, L. J., Bashem, J. R., & Hanks, R. A. (2019). Detecting malingering in traumatic brain injury: Combining response time with performance validity test accuracy. *The Clinical Neuropsychologist*, 33(1), 90–107. <https://doi.org/10.1080/13854046.2018.1440006>
- Kashluba, S., Paniak, C., Blake, T., Reynolds, S., Toller-Lobe, G., & Nagy, J. (2004). A longitudinal, controlled study of patient complaints following treated mild traumatic brain injury. *Archives of Clinical Neuropsychology*, 19(6), 805–816. <https://doi.org/10.1016/j.acn.2003.09.005>
- Kaufmann, P. M. (2012). Admissibility of expert opinions based on neuropsychological evidence. In G. J. Larrabee (Ed.) *Forensic Neuropsychology: A Scientific Approach* (2nd ed., pp. 70–100). Oxford University.
- Kemp, S., Coughlan, A. K., Rowbottom, C., Wilkinson, K., Teggart, V., & Baker, G. (2008). The base rate of effort test failure in patients with medically unexplained symptoms. *Journal of Psychosomatic Research*, 65(4), 319–325. <https://doi.org/10.1016/j.jpsychores.2008.02.010>
- Kirk, J. W., Baker, D. A., Kirk, J. J., & MacAllister, W. S. (2020). A review of performance and symptom validity testing with pediatric populations. *Applied Neuropsychology: Child*, 9(4), 292–306. <https://doi.org/10.1080/21622965.2020.1750118>
- Kirkwood, M. W. (2015a). *Validity testing in child and adolescent assessment: Evaluating exaggeration, feigning, and noncredible effort*. Guilford.
- Kirkwood, M. W. (2015b). Review of pediatric performance and symptom validity tests. In M. W. Kirkwood (Ed.), *Validity testing in child and adolescent assessment: Evaluating exaggeration, feigning, and noncredible effort*. Guilford.
- Kirkwood, M. W., Hargrave, D. D., & Kirk, J. W. (2011). The value of the WISC-IV Digit Span subtest in detecting noncredible performance during pediatric neuropsychological examinations. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 26(5), 377–384. <https://doi.org/10.1093/arclin/acr040>
- Kirkwood, M. W., & Kirk, J. W. (2010). The base rate of suboptimal effort in a pediatric mild TBI sample: Performance on the medical symptom validity test. *The Clinical Neuropsychologist*, 24(5), 860–872. <https://doi.org/10.1080/13854040903527287>
- Kizilbash, A. H., Vanderploeg, R. D., & Curtiss, G. (2002). The effects of depression and anxiety on memory performance. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 17(1), 57–67.
- Krahn, L. E., Bostwick, J. M., & Stonnington, C. M. (2008). Looking toward DSM-V: should factitious disorder become a subtype of somatoform disorder? *Psychosomatics*, 49(4), 277–282. <https://doi.org/10.1176/appi.psy.49.4.277>
- Kuelz, A. K., Hohagen, F., & Voderholzer, U. (2004). Neuropsychological performance in obsessive-compulsive disorder: A critical review. *Biological Psychology*, 65(3), 185–236. <https://doi.org/10.1016/j.biopsycho.2003.07.007>

- Lange, R. T., Brickell, T. A., French, L. M., Merritt, V. C., Bhagwat, A., Pancholi, S., & Iverson, G. L. (2012). Neuropsychological outcome from uncomplicated mild, complicated mild, and moderate traumatic brain injury in US military personnel. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 27(5), 480–494. <https://doi.org/10.1093/arclin/acs059>
- Lange, R. T., Iverson, G. L., & Rose, A. (2010). Post-concussion symptom reporting and the "good-old-days" bias following mild traumatic brain injury. *Archives of Clinical Neuropsychology*, 25(5), 442–450. <https://doi.org/10.1093/arclin/acq031>
- Lange, R. T., Lippa, S. M., Bailie, J. M., Wright, M., Driscoll, A., Sullivan, J., Gartner, R., Ramin, D., Robinson, G., Eshera, Y., Gillow, K., French, L. M., & Brickell, T. A. (2020). Longitudinal trajectories and risk factors for persistent postconcussion symptom reporting following uncomplicated mild traumatic brain injury in U.S. Military service members. *The Clinical Neuropsychologist*, 34(6), 1134–1155. <https://doi.org/10.1080/13854046.2020.1746832>
- Larrabee, G. J. (Ed.). (2007). *Assessment of malingered neuropsychological deficits*. Oxford University.
- Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, 17(3), 410–425. <https://doi.org/10.1076/clin.17.3.410.18089>
- Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist*, 22(4), 666–679. <https://doi.org/10.1080/13854040701494987>
- Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society: JINS*, 18(4), 625–630. <https://doi.org/10.1017/s1355617712000240>
- Larrabee, G. J. (2014). False-positive rates associated with the use of multiple performance and symptom validity tests. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 29(4), 364–373. <https://doi.org/10.1093/arclin/acu019>
- Larrabee, G. J., Bianchini, K. J., Boone, K. B., & Rohling, M. L. (2017). The validity of the MMPI-2/MMPI-2-RF symptom validity scale (FBS/FBS-r) is established: Reply to Nichols (2017). *The Clinical Neuropsychologist*, 31(8), 1401–1405. <https://doi.org/10.1080/13854046.2017.1363293>
- Larrabee, G. J., Greiffenstein, M. F., Greve, K. W., & Bianchini, K. J. (2007). Refining diagnostic criteria for malingering. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 334–371). Oxford University.
- Larrabee, G. J., Rohling, M. L., & Meyers, J. E. (2019). Use of multiple performance and symptom validity measures: Determining the optimal per test cutoff for determination of invalidity, analysis of skew, and inter-test correlations in valid and invalid performance groups. *The Clinical Neuropsychologist*, 33(8), 1354–1372. <https://doi.org/10.1080/13854046.2019.1614227>
- Lippa, S. M. (2018). Performance validity testing in neuropsychology: a clinical guide, critical review, and update on a rapidly evolving literature. *The Clinical Neuropsychologist*, 32(3), 391–421. <https://doi.org/10.1080/13854046.2017.1406146>
- Locke, D., Denham, M. E., Williamson, D. J., & Drane, D. (2017). Three faces of PNES. In K.B. Boone (Ed.), *Neuropsychological evaluation of somatoform and other functional somatic conditions: Assessment primer* (pp. 147–172). Routledge/Taylor & Francis.
- Lockhart, J., & Satya-Murti, S. (2015). Symptom exaggeration and symptom validity testing in persons with medically unexplained neurologic presentations. *Neurology: Clinical Practice*, 5, 17–24.
- Lopez, G., Ruiz, N., Patten, E. (2017). Key facts about Asian Americans, a diverse and growing population. Retrieved on March 6, 2020 from <https://www.pewresearch.org/fact-tank/2017/09/08/key-facts-about-asian-americans/>
- Loring, D. W., Goldstein, F. C., Chen, C., Drane, D. L., Lah, J. J., Zhao, L., & Larrabee, G. J. (2016). False-positive error rates for Reliable Digit Span and Auditory Verbal Learning Test performance validity measures in amnesic mild cognitive impairment and early Alzheimer Disease. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 31(4), 313–331. <https://doi.org/10.1093/arclin/acw014>

- Maffini, C. S., & Wong, Y. J. (2014). Assessing somatization with Asian American clients. In *Guide to psychological assessment with Asians*. (pp. 347–360). Springer.
- Marcopulos, B. A. (2018). Neuropsychological functioning in affective and anxiety-spectrum disorders in adults and children. In J.E. Morgan & J.H. Ricker (Eds.), *Textbook of clinical neuropsychology* (2nd ed., pp. 701–716). Routledge/Taylor & Francis.
- Marion, B. E., Sellbom, M., & Bagby, R. M. (2011). The detection of feigned psychiatric disorders using the MMPI-2-RF overreporting validity scales: An analog investigation. *Psychological Injury and Law*, 4(1), 1–12. <https://doi.org/10.1007/s12207-011-9097-0>
- Markova, H., Andel, R., Stepankova, H., Kopecek, M., Nikolai, T., Hort, J., Thomas-Antérion, C., & Vyhnaek, M. (2017). Subjective cognitive complaints in cognitively healthy older adults and their relationship to cognitive performance and depressive symptoms. *Journal of Alzheimer's Disease*, 59(3), 871–881. <https://doi.org/10.3233/JAD-160970>
- Marshall, P., Schroeder, R., O'Brien, J., Fischer, R., Ries, A., Blesi, B., & Barker, J. (2010). Effectiveness of symptom validity measures in identifying cognitive and behavioral symptom exaggeration in adult attention deficit hyperactivity disorder. *The Clinical Neuropsychologist*, 24(7), 1204–1237.
- Marshall, P. S., & Schroeder, R. W. (2021). Validity assessment in patients with psychiatric disorders. In R. W. Schroeder & P. K. Martin (Eds.), *Validity assessment in clinical neuropsychological practice*. Guilford.
- Martin, P. K., & Schroeder, R. W. (2020). Base rates of invalid test performance across clinical non-forensic contexts and settings. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 35(6), 717–725. <https://doi.org/10.1093/arclin/aca017>
- Martin, P. K., Schroeder, R. W., Heinrichs, R. J., & Baade, L. E. (2015). Does true neurocognitive dysfunction contribute to Minnesota multiphasic personality inventory-2nd edition-restructured form cognitive validity scale scores? *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 30(5), 377–386. <https://doi.org/10.1093/arclin/acv032>
- Martin, P., K., Schroeder, R. W., & Odland, A. P. (2015). Neuropsychologists' Validity Testing Beliefs and Practices: A Survey of North American Professionals. *The Clinical Neuropsychologist*, 29(6), 741–776. <https://doi.org/10.1080/13854046.2015.1087597>
- Martin, P. K., Schroeder, R. W., & Olsen, D. H. (2020). Performance validity in the dementia clinic: Specificity of validity tests when used individually and in aggregate across levels of cognitive impairment severity. *The Clinical Neuropsychologist*, <https://doi.org/10.1080/13854046.2020.1778790>
- Martin, P. K., Schroeder, R., W., Olsen, D. H., Maloy, H., Boettcher, A., Ernst, N., & Okut, H. (2020). A systematic review and meta-analysis of the Test of Memory Malingering in adults: Two decades of deception detection. *The Clinical Neuropsychologist*, 34(1), 88–119. <https://doi.org/10.1080/13854046.2019.1637027>
- Matto, M., McNeil, D. E., & Binder, R. L. (2019). A systematic approach to the detection of false PTSD. *The Journal of the American Academy of Psychiatry and the Law*, 47(3), 325–334. <https://doi.org/10.29158/JAAPL.003853-19>
- McCaffrey, R. J., & Yantz, C. L. (2007). Cognitive complaints in multiple chemical sensitivity and toxic mold syndrome. In K.B. Boone (Ed.) *Assessment of feigned cognitive impairment: A Neuropsychological perspective*. (pp. 384–404). Guilford.
- McCaul, C., Boone, K. B., Ermshar, A., Cottingham, M., Victor, T. L., Ziegler, E., Zeller, M. A., & Wright, M. (2018). Cross-validation of the Dot Counting Test in a large sample of credible and noncredible patients referred for neuropsychological testing. *The Clinical Neuropsychologist*, 32(6), 1054–1067. <https://doi.org/10.1080/13854046.2018.1425481>
- McCrea, M. (2008). *Mild traumatic brain injury and post-concussion syndrome: The new evidence base for diagnosis and treatment*. NY: Oxford University.
- McCullumsmith, C. B., & Ford, C. V. (2011). Simulated illness: The factitious disorders and malingering. *Psychiatric Clinics of North America*, 34(3), 621–641. <https://doi.org/10.1016/j.psc.2011.05.013>

- Merckelbach, H., & Smith, G. P. (2003). Diagnostic accuracy of the Structured Inventory of Malingered Symptomatology (SIMS) in detecting instructed malingering. *Archives of Clinical Neuropsychology*, 18(2), 145–152.
- Merten, T., & Merckelbach, H. (2013). Introduction to malingering research and symptom validity assessment. *Journal of Experimental Psychopathology*, 4(1), 3–5. <https://doi.org/10.1177/204380871300400102>
- Meyers, J. E., & Volbrecht, M. E. (2003). A validation of multiple malingering detection methods in a large clinical sample. *Archives of Clinical Neuropsychology*, 18(3), 261–276. <https://doi.org/10.1093/arclin/18.3.261>
- Mickevičiene, D., Schrader, H., Obelieniene, D., Surkiene, D., Kunickas, R., Stovner, L. J., & Sand, T. (2004). A controlled prospective inception cohort study on the post-concussion syndrome outside the medicolegal context. *European Journal of Neurology*, 11(6), 411–419. <https://doi.org/10.1111/j.1468-1331.2004.00816.x>
- Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, 24(8), 1094–1102. <https://doi.org/10.1076/jcen.24.8.1094.8379>
- Morgan, J. E., Millis, S. R., & Mesnik, J. (2009). Malingered dementia and feigned psychosis. In J. E. Morgan & J. J. Sweet (Eds.), *Neuropsychology of malingering casebook*. (pp. 231–243). Routledge/Taylor & Francis.
- Morgan, J. E., & Sweet, J. J. (2008). *Neuropsychology of malingering casebook*. Psychology Press/Taylor & Francis.
- Murrough, J. W., Iacoviello, B., Neumeister, A., Charney, D. S., & Iosifescu, D. V. (2011). Cognitive dysfunction in depression: Neurocircuitry and new therapeutic strategies. *Neurobiology of Learning and Memory*, 96(4), 553–563. <https://doi.org/10.1016/j.nlm.2011.06.006>
- Mutchnick, M. G., & Williams, J. M. (2012). Anxiety and memory test performance. *Applied Neuropsychology: Adult*, 19(4), 241–248. <https://doi.org/10.1080/09084282.2011.643965>
- Nelson, N. W., Hoelzle, J. B., Sweet, J. J., Arbisi, P. A., & Demakis, G. J. (2010). Updated meta-analysis of the MMPI-2 Fake Bad Scale: Verified utility in forensic practice. *The Clinical Neuropsychologist*, 24(4), 701–724. <https://doi.org/10.1080/13854040903482863>
- Nelson, N. W., Sweet, J. J., Berry, D. T. R., Bryant, F. B., & Granacher, R. P. (2007). Response validity in forensic neuropsychology: Exploratory factor analytic evidence of distinct cognitive and psychological constructs. *Journal of the International Neuropsychological Society: JINS*, 13(3), 440–449. <https://doi.org/10.1017/S1355617707070373>
- Nguyen, H.-H D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *The Journal of Applied Psychology*, 93(6), 1314–1334. <https://doi.org/10.1037/a0012702>
- Nijdam-Jones, A., & Rosenfeld, B. (2017). Cross-cultural feigning assessment: A systematic review of feigning instruments used with linguistically, ethnically, and culturally diverse samples. *Psychological Assessment*, 29(11), 1321–1336. <https://doi.org/10.1037/pas0000438>
- Ponsford, J., Cameron, P., Fitzgerald, M., Grant, M., Mikocka-Walus, A., & Schönberger, M. (2012). Predictors of postconcussive symptoms 3 months after mild traumatic brain injury. *Neuropsychology*, 26(3), 304–313. <https://doi.org/10.1037/a0027888>
- Proto, D. A., Pastorek, N. J., Miller, B. I., Romesser, J. M., Sim, A. H., & Linck, J. F. (2014). The dangers of failing one or more performance validity tests in individuals claiming mild traumatic brain injury-related postconcussive symptoms. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 29(7), 614–624. <https://doi.org/10.1093/arclin/acu044>
- Ray, C. L. (2014). Feigning screeners in VA PTSD compensation and pension examinations. *Psychological Injury and Law*, 7(4), 370–387. <https://doi.org/10.1007/s12207-014-9210-2>
- Resnick, P. J., & Knoll, J. L. IV. (2018). Malingered psychosis. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (4th ed., pp. 98–121). Guilford.
- Rey, A. (1964). *L'examen Clinique en psychologie*. Universitaires de France.
- Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's *d*, and *r*. *Law and Human Behavior*, 29(5), 615–620. <https://doi.org/10.1007/s10979-005-6832-7>

- Richards, P. M., & Tussey, C. M. (2013). The neuropsychologist as expert witness: Testimony in civil and criminal settings. *Psychological Injury and Law*, 6(1), 63–74. <https://doi.org/10.1007/s12207-013-9148-9>
- Robles, L., López, E., Salazar, X., Boone, K. B., & Glaser, D. F. (2015). Specificity data for the *b* Test, Dot Counting Test, Rey-15 Item Plus Recognition, and Rey Word Recognition Test in monolingual Spanish-speakers. *Journal of Clinical and Experimental Neuropsychology*, 37(6), 614–621. <https://doi.org/10.1080/13803395.2015.1039961>
- Rogers, R., Sewell, K. W., & Gillard, N. D. (2010). *SIRS-2: Structured interview of reported symptoms: Professional manual*. Psychological Assessment Resources.
- Rohling, M. L., Binder, L. M., Demakis, G. J., Larrabee, G. J., Ploetz, D. M., & Langhinrichsen-Rohling, J. (2011). A meta-analysis of neuropsychological outcome after mild traumatic brain injury: Re-analyses and reconsiderations of Binder et al. (1997), Frencham et al. (2005), and Pertab et al. (2009). *The Clinical Neuropsychologist*, 25(4), 608–623. <https://doi.org/10.1080/13854046.2011.565076>
- Rohling, M. L., Langhinrichsen-Rohling, J., & Meyers, J. E. (2021). Effects of Premorbid Ability, Neuropsychological Impairment and Invalid Test Performance on the Frequency of Low Scores (Chapter 13). In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (2nd ed.). Guilford.
- Rohling, M. L., Langhinrichsen-Rohling, J., & Womble, M. N. (2015). Pediatric sports-related concussion evaluations. In M. W. Kirkwood (Ed.), *Validity testing in child and adolescent assessment* (pp. 226–249). Guilford.
- Rohling, M. L., Meyers, J. E., & Millis, S. R. (2003). Neuropsychological impairment following TBI: A dose response analysis. *The Clinical Neuropsychologist*, 17(3), 289–302. <https://doi.org/10.1076/clin.17.3.289.18086>
- Rosner, R., & Scott, C. (Eds.). (2016). *Principles and practice of forensic psychiatry* (3rd ed.). CRC Press/Taylor & Francis.
- Ruff, R. M., Camenzuli, L., & Mueller, J. (1996). Miserable minority: Emotional risk factors that influence the outcome of a mild traumatic brain injury. *Brain Injury*, 10(8), 551–566. <https://doi.org/10.1080/026990596124124>
- Ruiz, I., Raugh, I. M., Bartolomeo, L. A., & Strauss, G. P. (2020). A meta-analysis of neuropsychological effort test performance in psychotic disorders. *Neuropsychology Review*, 30(3), 407–418. <https://doi.org/10.1007/s11065-020-09448-2>
- Salazar, X., Lu, P., & Boone, K. B. (2021). The Use of Performance Validity Tests in Ethnic Minorities and Non-English Dominant Populations. In Boone, K. B. (Ed.), *Assessment of Feigned Cognitive Impairment: A Neuropsychological Perspective*. (2nd ed.). Guilford.
- Scamvougeras, A., & Howard, A. (2018). *Understanding and managing somatoform disorders: A guide for clinicians*. Vancouver., AJKS Medical.
- Schroeder, R. W., & Martin, P. K. (2021a). Explanations of performance validity test failure in clinical settings. In R. W. Schroeder & P. K. Martin (Eds.), *Validity assessment in clinical neuropsychological practice*. Guilford. (in press).
- Schroeder, R. W., & Martin, P. K. (2021b). Validity assessment in clinical settings: How it differs from forensic settings and why it is important. In R. W. Schroeder & P. K. Martin (Eds.), *Validity assessment in clinical neuropsychological practice*. Guilford. (in press).
- Schroeder, R. W., & Martin, P. K. (2021c). Validity assessment within the memory disorders/dementia clinic. In R. W. Schroeder & P. K. Martin (Eds.), *Validity assessment in clinical neuropsychological practice*. Guilford. (in press).
- Schroeder, R. W., & Martin, P. K. (Eds.) (2021d). *Validity assessment in clinical neuropsychological practice*. Guilford. (in press).
- Schroeder, R. W., Martin, P. K., Heinrichs, R. J., & Baade, L. E. (2019). Research methods in performance validity testing studies: Criterion grouping approach impacts study outcomes. *The Clinical Neuropsychologist*, 33(3), 466–477. <https://doi.org/10.1080/13854046.2018.1484517>
- Schroeder, R. W., Martin, P. K., & Odland, A. P. (2016). Expert beliefs and practices regarding neuropsychological validity testing. *The Clinical Neuropsychologist*, 30(4), 515–535. <https://doi.org/10.1080/13854046.2016.1177118>

- Schroeder, R. W., Olsen, D. H., & Martin, P. K. (2019). Classification accuracy rates of four TOMM validity indices when examined independently and conjointly. *The Clinical Neuropsychologist*, 33(8), 1373–1387. <https://doi.org/10.1080/13854046.2019.1619839>
- Sellbom, M., & Bagby, R. M. (2010). Detection of overreported psychopathology with the MMPI-2-RF [corrected] validity scales. *Psychological Assessment*, 22(4), 757–767. <https://doi.org/10.1037/a0020825>
- Sharf, A. J., Rogers, R., Williams, M. M., & Henry, S. A. (2017). The effectiveness of the MMPI-2-RF in detecting feigned mental disorders and cognitive deficits: A meta-analysis. *Journal of Psychopathology and Behavioral Assessment*, 39(3), 441–455. <https://doi.org/10.1007/s10862-017-9590-1>
- Shaw, R. J., Dayal, S., Hartman, J. K., & Demaso, D. R. (2008). Factitious disorder by proxy: Pediatric condition falsification. *Harvard Review of Psychiatry*, 16(4), 215–224. <https://doi.org/10.1080/10673220802277870>
- Sherman, E. M. S., Slick, D. J., & Iverson, G. L. (2020). Multidimensional malingering criteria for neuropsychological assessment: A 20-year update of the malingered neuropsychological dysfunction criteria. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 35(6), 735–764. <https://doi.org/10.1093/arclin/acia019>
- Sjöwall, D., Roth, L., Lindqvist, S., & Thorell, L. B. (2013). Multiple deficits in ADHD: Executive dysfunction, delay aversion, reaction time variability, and emotional deficits. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 54(6), 619–627. <https://doi.org/10.1111/jcpp.12006>
- Slick, D. J., & Sherman, E. M. S. (2012). Differential diagnosis of malingering and related clinical presentations. In E. M. Sherman & B. L. Brooks (Eds.), *Pediatric forensic neuropsychology* (pp. 113–135). Guilford.
- Slick, D. J., & Sherman, E. M. S. (2013). Differential diagnosis of malingering. In D.A. Carone & S.S. Bush (Eds.), *Mild traumatic brain injury: Symptom validity assessment and malingering* (pp. 57–72). Springer.
- Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, 13(4), 545–561. [https://doi.org/10.1076/1385-4046\(199911\)13:04;1-Y;FT545](https://doi.org/10.1076/1385-4046(199911)13:04;1-Y;FT545)
- Sobanski, E., Banaschewski, T., Asherson, P., Buitelaar, J., Chen, W., Franke, B., Holtmann, M., Krumm, B., Sergeant, J., Sonuga-Barke, E., Stringaris, A., Taylor, E., Anney, R., Ebstein, R. P., Gill, M., Miranda, A., Mulas, F., Oades, R. D., Roeyers, H., ... Faraone, S. V. (2010). Emotional lability in children and adolescents with attention deficit/hyperactivity disorder (ADHD): Clinical correlates and familial prevalence. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 51(8), 915–923. <https://doi.org/10.1111/j.1469-7610.2010.02217.x>
- Sollman, M. J., & Berry, D. T. R. (2011). Detection of inadequate effort on neuropsychological testing: A meta-analytic update and extension. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 26(8), 774–789. <https://doi.org/10.1093/arclin/acr066>
- Suchy, Y. (2019). Introduction to special issue: Current trends in empirical examinations of performance and symptom validity. *The Clinical Neuropsychologist*, 33(8), 1349–1343. <https://doi.org/10.1080/13854046.2019.1672334>
- Suhr, J. A., Cook, C., & Morgan, B. (2017). Assessing functional impairment in ADHD: Concerns for validity of self-report. *Psychological Injury and Law*, 10(2), 151–160. <https://doi.org/10.1007/s12207-017-9292-8>
- Suhr, J. A., & Gunstad, J. (2002). "Diagnosis Threat": the effect of negative expectations on cognitive performance in head injury". *Journal of Clinical and Experimental Neuropsychology*, 24(4), 448–457. <https://doi.org/10.1076/jcen.24.4.448.1039>
- Suhr, J., & Spickard, B. (2007). Including measures of effort in neuropsychological assessment of pain- and fatigue-related medical disorders: Clinical and research implications. In K.B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 259–280). Guilford.

- Suhr, J., & Wei, C. (2017). Attention deficit/hyperactivity disorder as an illness identity: Implications for neuropsychological practice. In K. B. Boone (Ed.), *Neuropsychological evaluation of somatoform and other functional somatic conditions: Assessment primer* (pp. 251–273). Routledge/Taylor & Francis.
- Sweet, J. J., Benson, L. M., Nelson, N. W., & Moberg, P. J. (2015). The American Academy of Clinical Neuropsychology, National Academy of Neuropsychology, and Society for Clinical Neuropsychology (APA Division 40) 2015 TCN professional practice and 'Salary Survey': Professional Practices, Beliefs, and Incomes of U.S. Neuropsychologists. *The Clinical Neuropsychologist*, 29(8), 1069–1162. <https://doi.org/10.1080/13854046.2016.1140228>
- Sweet, J. J., Goldman, D. J., & Guidotti Breting, L. M. (2013). Traumatic brain injury: Guidance in a forensic context from outcome, dose-response, and response bias research. *Behavioral Sciences & the Law*, 31(6), 756–778. <https://doi.org/10.1002/bsl.2088>
- Sweet, J. J., & Guidotti Breting, L. M. (2013). Symptom validity test research: Status and clinical implications. *Journal of Experimental Psychopathology*, 4(1), 6–19. <https://doi.org/10.5127/jep.022311>
- Sweet, J. J., Kaufmann, P. M., Ecklund-Johnson, E., & Malina, A. C. (2018). Forensic neuropsychology: An overview of issues, admissibility, and directions. In J. E. Morgan & J. H. Ricker (Eds.) *Textbook of clinical neuropsychology* (2nd ed., pp. 857–886). Routledge/Taylor & Francis.
- Sweet, J. J., Klipfel, K. M., Nelson, N. W., & Moberg, P. J. (2021a). Professional Practices, Beliefs, and Incomes of postdoctoral trainees: The AACN, NAN, SCN 2020 Practice and 'Salary Survey'. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 36(1), 1–16. <https://doi.org/10.1093/arclin/acia116>
- Sweet, J. J., Klipfel, K. M., Nelson, N. W., & Moberg, P. J. (2021b). Professional Practices, Beliefs, and Incomes of U.S. Neuropsychologists: The AACN, NAN, SCN 2020 Practice and "salary survey". *The Clinical Neuropsychologist*, 35(1), 7–80. <https://doi.org/10.1080/13854046.2020.1849803>
- Tombaugh, T. N. (1996). *Test of Memory Malingering: TOMM*. Multi-Health Systems.
- Tylicki, J. L., Gervais, R. O., & Ben-Porath, Y. S. (2020). Examination of the MMPI-3 over-reporting scales in a forensic disability sample. *The Clinical Neuropsychologist*, <https://doi.org/10.1080/13854046.2020.1856414>
- Van Dyke, S. A., Millis, S. R., Axelrod, B. N., & Hanks, R. A. (2013). Assessing effort: Differentiating performance and symptom validity. *The Clinical Neuropsychologist*, 27(8), 1234–1246. <https://doi.org/10.1080/13854046.2013.835447>
- Vanderploeg, R. D., Belanger, H. G., & Kaufmann, P. M. (2014). Nocebo effects and mild traumatic brain injury: Legal implications. *Psychological Injury and Law*, 7(3), 245–254. <https://doi.org/10.1007/s12207-014-9201-3>
- Vickery, C. D., Berry, D. T. R., Inman, T. H., Harris, M. J., & Orey, S. A. (2001). Detection of inadequate effort on neuropsychological testing: A meta-analytic review of selected procedures. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 16(1), 45–73.
- Victor, T. L., Boone, K. B., Serpa, J. G., Buehler, J., & Ziegler, E. A. (2009). Interpreting the meaning of multiple symptom validity test failure. *The Clinical Neuropsychologist*, 23(2), 297–313. <https://doi.org/10.1080/13854040802232682>
- Webber, T. A., & Soble, J. R. (2018). Utility of various WAIS-IV Digit Span indices for identifying noncredible performance validity among cognitively impaired and unimpaired examinees. *The Clinical Neuropsychologist*, 32(4), 657–670. <https://doi.org/10.1080/13854046.2017.1415374>
- Wehmeier, P. M., Schacht, A., & Barkley, R. A. (2010). Social and emotional impairment in children and adolescents with ADHD and the impact on quality of life. *The Journal of Adolescent Health: official Publication of the Society for Adolescent Medicine*, 46(3), 209–217. <https://doi.org/10.1016/j.jadohealth.2009.09.009>
- Whitman, M. R., Tylicki, J. L., Mascioli, R., Pickle, J., & Ben-Porath, Y. S. (2020). Psychometric Properties of the Minnesota Multiphasic Personality Inventory-3 (MMPI-3) in a clinical neuropsychology setting. *Psychological Assessment*, <https://doi.org/10.1037/pas0000969>

- Williamson, K. D., Combs, H. L., Berry, D. T., Harp, J. P., Mason, L. H., & Edmundson, M. (2014). Discriminating among ADHD alone, ADHD with a comorbid psychological disorder, and feigned ADHD in a college sample. *The Clinical Neuropsychologist*, 28(7), 1182–1196. <https://doi.org/10.1080/13854046.2014.956674>
- Williamson, D. J., Ogden, M. L., & Drane, D. L. (2015). Psychogenic non-epileptic seizures (PNES): an update. In S. P. Koffler, J. E. Morgan, B. A. Marcopulos (Eds.), *Neuropsychology: Science and practice* (pp. 1–25). Oxford University.
- Wood, R. L., O'Hagan, G., Williams, C., McCabe, M., & Chadwick, N. (2014). Anxiety sensitivity and alexithymia as mediators of postconcussion syndrome following mild traumatic brain injury. *Journal of Head Trauma Rehabilitation*, 29(1), E9–E17. <https://doi.org/10.1097/HTR.0b013e31827eabba>
- Wygant, D. B., Ben-Porath, Y. S., Arbisi, P. A., Berry, D. T., Freeman, D. B., & Heilbronner, R. L. (2009). Examination of the MMPI-2 restructured form (MMPI-2-RF) validity scales in civil forensic settings: Findings from simulation and known group samples. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 24(7), 671–680. <https://doi.org/10.1093/arclin/acp073>
- Wygant, D. B., Sellbom, M., Ben-Porath, Y. S., Stafford, K. P., Freeman, D. B., & Heilbronner, R. L. (2007). The relation between symptom validity testing and MMPI-2 scores as a function of forensic evaluation context. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 22(4), 489–499. <https://doi.org/10.1016/j.acn.2007.01.027>
- Zhou, X.-H., Obuchowski, N. A., & McClish, D. K. (2002). *Statistical methods in diagnostic medicine*. Wiley-Interscience.